Legal Theory, **19** (2013), 403–421.
© Cambridge University Press 2014 0361-6843/14 $15.00 + 00
doi:10.1017/S1352325213000219

LEGALITY AND RATIONALITY: A COMMENT ON SCOTT SHAPIRO'S *LEGALITY*

Emily Sherwin*
Cornell Law School

One key premise in Shapiro's book *Legality* is that rationality requires those who have accepted the master plan for a system of law to obey the system's rules. In this paper, I question this premise, arguing instead that although it may be rational for agents to commit to follow the system's rule in all (or most) cases to which they apply, it is not rational for agents to follow the rules in fact when the rules appear to require the wrong outcomes in particular cases. My argument is based largely on epistemic responsibility, which I view as an element of rationality.

In his book *Legality*, Scott Shapiro depicts law as a system of interconnected plans for ordering conduct in a complex society.[1] Legal activity—the making and application of laws—is a process of social planning that aims to settle moral controversy. The capacity of lawmakers to establish legal plans, as well as the binding effect of legal plans on those who execute, enforce, and act under law, is governed by principles of instrumental rationality. These principles dictate, among other things, that those who accept legal plans are bound to comply with them.

Shapiro's theory of law is a positivist theory: the status of particular plans as laws does not depend on their conformity to moral standards. Instead, legality depends on the social fact that legal actors have accepted a master plan providing for institutional development of a system of shared, compulsory plans.[2] Legal rules established according to the master plan have normative force but only from a legal point of view.[3] That is, laws *purport* to impose moral obligations on agents and they do in fact impose moral obligations on those who accept the master plan. Those who do not accept the plan may be coerced to comply but have no moral obligation to do so.

*Thanks to Larry Alexander and to participants in a faculty workshop at Cornell for helpful comments.

1. SCOTT J. SHAPIRO, LEGALITY (2011). In characterizing laws as plans, Shapiro draws on Michael Bratman's work on the function of planning in practical reason. *See* MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON (1987).

2. Shapiro's master plan appears to be a set of "fundamental legal rules" that authorizes some members of the community to plan for others. *Id.* at 180–183.

3. *Id.* at 184–188.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

The system of laws also has a moral aim—to better the condition of the community by settling a range of moral controversies.[4] Yet a legal system that fails to accomplish settlement or accomplishes settlement in morally unacceptable ways is nevertheless a system of law.[5]

Shapiro develops his conception of law in the first half of the book; later chapters are devoted to implications of the thesis, especially in the area of interpretation. Over the course of the book, Shapiro steers a steady and logical path through the difficult terrain of jurisprudence. Fortunately for his readers, he is a master of explanation. He describes with meticulous care and unusual clarity the positivist theories of John Austin and H.L.A. Hart as well as the challenges posed by Ronald Dworkin and others who reject legal positivism. Even mysteries such as the difference between inclusive legal positivism and exclusive legal positivism appear suddenly (if only momentarily) clear. Shapiro applies the same explanatory skills in laying out his own planning theory of law as well as a related theory of "meta-interpretation."[6]

My focus in this essay is on the question of rationality and the binding effect of law (or legal plans). Shapiro writes with confidence about the rationality of participation in and obedience to law.[7] In a book that covers many of the same themes that Shapiro addresses in *Legality*, Larry Alexander and I argue to the contrary, that it is not rational to follow the law in the unvarying manner that law requires.[8] Instead, there is an inescapable gap between the level of compliance that rational lawmakers have reason to demand and the instances in which rational agents acting under the law have reason to comply. As a result, law cannot be fully effective as a means of social ordering and moral settlement without the help of some form of widespread irrationality, such as popular self-deception about the reasons for action that law provides.[9]

My aim here is to account for the difference between Shapiro's view and ours and to defend our skeptical conclusion. In my view, Shapiro errs in isolating instrumental rationality from other aspects of practical deliberation. I propose that although there may be instrumental reasons to hold that acceptance of a rule or plan generates a binding commitment to comply with the plan, instrumental rationality is not the only standard that governs compliance with law.[10] Rationality may also encompass or at least depend on

4. *Id.* at 213–217.
5. *Id.* at 187–188.
6. *Id.* at 355.
7. *Id.* at 183.
8. LARRY ALEXANDER & EMILY SHERWIN, THE RULE OF RULES: MORALITY, RULES, AND THE DILEMMAS OF LAW (2001).
9. *Id.* at 53–95.
10. For a discussion of the relationship between instrumental rationality and epistemic rationality, see, e.g., Thomas Kelly, *Evidence and Normativity: Reply to Leite*, 75 PHILOSOPHY & PHENOMENOLOGICAL RESEARCH 165–174 (2007). For an argument that instrumental rationality does not exhaust the concept of rationality and indeed is not a distinctive form of rationality at all, *see* Joseph Raz, *The Myth of Instrumental Rationality*, 1 J. ETHICS & SOC. PHIL. 1 (2005). Raz is not specifically concerned with epistemic standards.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

standards of epistemic responsibility. When an agent must decide whether or not to follow the law in a particular case, epistemic responsibility can diverge from instrumental rationality and militate against compliance. In cases of divergence, epistemic responsibility is naturally prior to instrumental rationality. Taking epistemic responsibility into account, I hope to show that an agent can (rationally) endorse a legal plan and yet be unable (rationally) to comply with the plan when the time for compliance arrives.

I. SOME SHARED PREMISES

Alexander and I begin from premises that are strikingly similar to the premises Shapiro sets out in *Legality*. What we call the "Circumstances of Law"[11] are much the same as Shapiro's "Circumstances of Legality."[12] We describe a community whose members agree on the general principles of morality but disagree about specific applications of those principles;[13] likewise, Shapiro imagines a community of morally virtuous persons who nevertheless face moral "doubts and disagreements" that cannot effectively be resolved by nonlegal means.[14] Accordingly, the fundamental aim of law (or legal planning) is the same for us and for Shapiro: our lawmaker, Lex, is chosen by general accord to provide authoritative settlement of moral controversy;[15] Shapiro's lawmaker, also called Lex, is vested by general accord with authority to make plans that "rectify the moral deficiencies associated with the circumstances of legality."[16] Authoritative settlement, we argue, improves the moral prospects of the community by coordinating individual conduct, providing expertise, and increasing the efficiency of moral decision-making;[17] Shapiro, too, relies on the need for coordination and efficiency as well as "an ethic of mutual respect" arising from shared agency.[18]

Alexander and I identify the principal means for settling moral controversy within a society as legal rules; Shapiro helpfully redescribes laws as plans—a term that captures the cumulative, interconnected, and open-ended nature of a system of laws.[19] The essential characteristics we ascribe to rules, however, are very similar to the characteristics Shapiro ascribes to legal plans. In particular, we maintain that effective legal rules must be

11. ALEXANDER & SHERWIN, *supra* note 8, at 9.
12. SHAPIRO, *supra* note 1, at 170. *Cf.* DAVID HUME, AN ENQUIRY CONCERNING THE PRINCIPLES OF MORALS (1751), pt. 1, sec. 3; DAVID HUME, 3 A TREATISE OF HUMAN NATURE (1739–1740), pt. 2, sec. 2.
13. ALEXANDER & SHERWIN, *supra* note 8, at 11–12.
14. SHAPIRO, *supra* note 1, at 170.
15. ALEXANDER & SHERWIN, *supra* note 8, at 15–17.
16. SHAPIRO, *supra* note 1, at 36–40, 180–181, 213.
17. ALEXANDER & SHERWIN, *supra* note 8, at 14.
18. SHAPIRO, *supra* note 1, at 150–153.
19. Continuing the parallel, we title our book "The Rule of Rules;" Shapiro refers in his concluding chapter to "The Rule of Plans." ALEXANDER & SHERWIN, *supra* note 8; SHAPIRO, *supra* note 1, at 392.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

general in application and also reasonably determinate in form. Addressing determinacy, we say that "if uncertainty and disagreement about how moral principles apply in concrete situations are what give rise to the need for Lex and Lex's rules, Lex's rules are responsive to this need only if they themselves do not engender uncertainty and disagreement about their concrete applications."[20] Shapiro makes the point perhaps more elegantly in the form of a principle he calls the "Simple Logic of Planning" ("SLOP"): if a legal plan is to settle moral controversy, "[t]he existence and content of a plan cannot be determined by facts whose existence the plan aims to settle."[21]

For Shapiro, SLOP prohibits moral content in legal plans. In contrast, we take the view that the use of moral or evaluative terms in legal rules is acceptable as long as the terms have an accepted meaning within the community.[22] Although this difference may have some theoretical significance, the underlying idea is the same: if rules or plans are to settle moral questions effectively, those who act under them must be able to follow their prescriptions without further deliberation on the questions the rules or plans are designed to settle.

Another point of agreement is that we, like Shapiro (and Hart), recognize the importance of a hierarchy of rules (or plans). A legal system must be founded on a master rule that exists by virtue of general recognition and renders the system authoritative. We also agree that an effective legal system needs secondary rules that order the process of making and applying rules and plans.[23]

Finally, Shapiro appears to share our sense of when a legal system is worthy of acceptance by those who are subject to its laws. I note above that rules (or plans) are designed to settle controversy by translating agreed moral principles into determinate prescriptions for action. Determinate prescriptions, if generally accepted and obeyed, can improve on unguided decision-making through the effects of coordination, decision-making efficiency, and, depending on the wisdom of the authority that issues them, comparative expertise. At the same time, the determinacy of rules makes them overinclusive and underinclusive in application: they will sometimes dictate outcomes that are incorrect when judged by the underlying principles on which they are based. Thus, in the course of reducing error, rules introduce errors of another kind.[24]

20. ALEXANDER & SHERWIN, *supra* note 8, at 32–33.
21. SHAPIRO, *supra* note 1, at 275.
22. ALEXANDER & SHERWIN, *supra* note 8, at 32–34.
23. *Id.* at 17–19, 37–49; SHAPIRO, *supra* note 1, at 165–168; 205–211. *See* H.L.A. HART, THE CONCEPT OF LAW (1964), at 91–94. Hart suggests that secondary rules of adjudication and change are part of what identifies law as law. For Alexander and me, secondary rules are desirable, but only the rule of recognition is necessary to accomplish settlement.
24. I use the term "error" here rather than "moral error" to avoid questions about the scope of morality. If pressed, I would define morality broadly and describe the effects of discord and controversy about moral questions as forms of moral error.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

This means that the desirability of rules (or plans) is contingent: they count as good rules worthy of acceptance when those who follow them consistently will commit fewer errors or errors of lesser magnitude than they would commit in the course of unguided practical reasoning. For ease of reference, I give this criterion for rule acceptance a name (the Net Error Reduction Principle) and an acronym (NERP).

> NERP: If an agent expects to commit less error by regularly following a rule R than the agent would commit by regularly engaging in unguided reasoning, instrumental rationality requires the agent to accept R as a rule to be followed in all cases to which it applies.

Notice that NERP requires the agent to accept R as a rule for all cases. If the agent were permitted to accept R provisionally, as a rule to be followed only when the agent judged that the outcome the rule calls for was correct, the rule would exercise no constraint on the agent's reasoning and yield no benefits. Thus the agent must endorse the rule as a rule to be followed in all cases that fall within its terms.

NERP also implies that the master plan of a legal system is worthy of acceptance when the overall effect of its present and future rules, taken together, is to reduce the level of error by agents. As a strict legal positivist, Shapiro would say that the question of whether a legal system succeeds in reducing the overall level of error has no effect on its status as a system of law. Nevertheless, as a criterion for external evaluation, including deliberation by potential participants in the system about whether to accept or reject the system's master plan, NERP is consistent with the primary moral aim of law.

II. DIFFERING CONCLUSIONS

Thus Alexander and I agree with Shapiro about the aim and benefits of law and, for the most part, about what is needed to obtain those benefits. At this point, however, our view of law diverges sharply from Shapiro's. For Shapiro, once a master plan for law wins general acceptance, the feasibility of legal planning falls into place.[25] The capacity to plan is part of human nature, and compliance with plans follows from the fact of acceptance. Those who accept must obey, because commitment to the system makes compliance the only rational option. Those who do not accept can be brought into line through sanctions for disobedience. The main challenge, as Shapiro sees it,

25. SHAPIRO, *supra* note 1, at 178–188. Shapiro does not provide extended discussion of the binding effect of plans in LEGALITY, but the assumption that plans can and do control future actions lies at the heart of his planning theory of law. Earlier work lays out the arguments for this assumption at greater length. *See* Scott Shapiro, *The Difference That Rules Make*, *in* ANALYZING LAW: NEW ESSAYS IN LEGAL THEORY (Brian Bix ed., 1998); Edward F. McClennen & Scott Shapiro, *Rule-Guided Behavior*, *in* 3 NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 363 (Peter Newman ed., 1998).

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

is how to interpret the plans made by various officials who play a part in the planning process. Shapiro proposes that interpretive methodology should vary according to the level of trust that those who accept the master plan repose in the relevant officials.[26]

In our view, the central problem in legal theory is how law can be effective in guiding action. Summarizing our position, we agree that it is rational for agents to endorse the system of legal rules (or legal planning) and by doing so to endorse all rules that issue from it, intending to follow them in all cases. But it does not follow for us that it is rational—let alone rationally compelled—for agents to conform to the rules in particular cases when those cases arise. Rules translate moral principles into concrete directions for classes of cases. They are accordingly different from and simpler than the moral principles that lie behind them and will sometimes dictate a morally wrong result. When agents consider rules in the abstract, the balance of reasons requires them to accept any rule that meets the test of NERP and to accept it as a rule to be followed in all cases. But when an agent must

26. *See* SHAPIRO, LEGALITY, *supra* note 1, at 331–387. Alexander and I also differ from Shapiro in our approach to interpretation. We believe that if the reason for having rules is to settle controversy, it must follow that the meaning of rules depends on the rule-maker's intent. When members of a community select an authority to settle controversy by means of rules, what they want is the set of rules that person chooses, after deliberation, to promulgate. The task of interpreters, therefore, is to determine what meaning the rule-making authority intended the words of the rule to have—not the rule-maker's purpose, not the standard meaning or best understanding of the rules' text, but the intended meaning of the rules. This is the only methodology we recognize as "interpretation." Intended meaning may not always be accessible; in fact, it may not always exist (as in the case of multimember authorities). When intent is not available, those who apply law can proceed in a number of ways, but none of these counts as interpretation. *See* ALEXANDER & SHERWIN, *supra* note 8, at 97–122; LARRY ALEXANDER & EMILY SHERWIN, DEMYSTIFYING LEGAL REASONING (2008), at 129–232. Given the settlement function Shapiro associates with plans, we would expect him to take a similar approach. Yet Shapiro equivocates. He appears to hold that a variety of so-called "interpretive" methods may be sound and that which of these a particular interpreter ought to use depends on how much trust the community places in the interpreter. But this is a confusion. In a system founded on settlement, the meaning of a determinate rule (or plan) must be the meaning that the person entrusted with settlement intended the rule to have; otherwise settlement does not occur at the level of planning but instead occurs at some point further down the line. If we are correct about this, then the community's trust in the interpreter is irrelevant. Trust is reposed in planners, not interpreters. In some cases, the legal norm may a relatively indeterminate norm—a standard or a plan that lays out standards for subsequent planning. In that case, the community may be trusting the person to whom the norm is addressed, but it is trusting that person to engage in further planning rather than to *interpret* the norm. In other words, interpretation ends when the addressee concludes that the norm in question is a standard. At that point, the norm is a delegation of authority, and the addressee must engage in first-order practical reasoning within the bounds set by the norm. In Shapiro's terms, the community is trusting the addressee to engage in subplanning. But subplanning and the trust reposed in subplanners have nothing to do with interpretation. Thus interpretation is always an attempt to recover the authorially intended meaning. If the legal planner chooses to settle matters by means of a determinate norm, then the objective is to determine what settlement the planner intended to impose. When the planner chooses a standard, the question is still one of intended meaning. In this case, the planner intends to delegate settlement to others. Once the delegatees discern this intention, they have no further interpretive role; their role is to plan.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

apply the rule to a set of facts, and the rule, so applied, prescribes an outcome that appears wrong according to the underlying principles the rule is designed to clarify and promote, the balance of reasons requires the agent to disobey. There is thus a gap between what the agent ought to endorse (an exceptionless rule) and what the agent ought to do (make an exception).

If agents were always correct in their assessment of what action shared moral values require in a particular case, this gap might not be of serious concern: an agent's decision not to follow the rule would be a moral improvement over the outcome indicated by the rule. But one of the principal reasons that rules (or plans) are necessary is that agents are often mistaken in their assessment of what to do, either because they act on insufficient evidence or because they are unable, without the aid of reliable rules, to coordinate their decisions with those of others. Therefore, it remains the case that the rule-maker ought to demand full compliance in all cases and that the agent ought to accept the rule on these terms.

A partial solution is what Frederick Schauer calls rule-sensitive particularism.[27] Rule-sensitive particularism treats the decision whether to follow a rule as an all-things-considered decision based on current evidence about reasons for action but counts the benefits of settlement as reasons to comply with the rule. At least in some circumstances, rule-sensitive particularism will lead an agent who might otherwise defect from the rule to follow it instead. Specifically, the agent may reason that his own defection will lead others who previously were disposed to follow the rule to reconsider their stance of unthinking obedience. The agent's disobedience, in other words, will set a potentially harmful example. If so, the others may decide to follow their judgment and disobey in cases in which their judgment is wrong. This prospect, in turn, becomes a reason for the original agent to obey the rule.

The difficulty is that under rule-sensitive particularism, the potential effect of a bad example is not a conclusive reason to obey: the agent may believe that the moral error the agent will commit by following the rule more than offsets any negative consequences of defection and so may defect after all. And again, the agent's calculation may be wrong. Thus rule-sensitive particularism fails to close the gap between the exceptionless rule the agent has reason to endorse and the action the agent has reason to take.[28]

27. Frederick Schauer, Playing by the Rules: A Philosophical Examination of Rule-Bound Decision-Making in Law and in Life (1991), at 94–100. McClennen and Shapiro refer to defenses of this type as "compatibilist" defenses: when the value of rules is property taken into account, the agent's reasons for action will be compatible with obedience to the rule. McClennen & Shapiro, *supra* note 25, at 364–365.

28. Alexander & Sherwin, Rule of Rules, *supra* note 8, at 68. We also examine and reject a variety of other putative solutions to the problem of the gap, including the possibility of sanctions. *Id.* at 68–95.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

III. RATIONALITY

Shapiro believes that those who accept a system of legal plans are practically committed to following the rules the plan imposes on them: rationality entails compliance.[29] Alexander and I believe that an agent can have undefeated reasons to accept a rule and commit to follow it in all future cases and yet also have undefeated reasons to disobey the rule in some of those cases when they arise. What explains this divergence?

Evidently we are working under different understandings of rationality. We can probably agree that rationality involves an appropriate response to applicable reasons for action or decision. For Alexander and me, this means an appropriate response to the full set of reasons that apply to the agent at the time the agent acts. Among possible acts—such as following a rule or not following a rule—it is rational for the agent to do what the weight of current reasons for action recommends.

Shapiro, in contrast, takes a view of rationality that is both broader and narrower than ours. It is broader in a temporal sense: rationality is a property of agents that cuts across time.[30] At the same time, Shapiro limits his attention to instrumental rationality: rationality requires agents to perform those actions that will best promote the agents' ends.[31] Applying Shapiro's temporally expansive notion of instrumental rationality to law, an agent may determine that given the fallibility of unguided reasoning, the best way to accomplish his or her ends is to commit to follow a legal plan and to carry through on the commitment in the future. In committing to follow the rule, the agent is "strategically interacting with another agent—his later self," with the terms of interaction fixed by the present self.[32] This means that from the time of commitment (T_C), the agent's decision is binding on future selves. The effect is to eliminate the option of future noncompliance: the whole self has committed and the future self has no independent choice in the matter. Departing from the rule's prescription is simply not "feasible."[33]

The term "feasible" may seem odd in this context because breaking a rule is surely something the agent can *do*. Shapiro, however, adopts a special notion of feasibility that he ties to reasons for action.[34] An action is feasible, in Shapiro's sense, if the agent can perform it *for a reason*; conversely, if the

29. *See supra* note 25 and accompanying text.
30. Shapiro, *Difference*, *supra* note 25, at 39, 44.
31. SHAPIRO, LEGALITY, *supra* note 1, at 181, 183; Shapiro, *Difference*, *supra* note 25, at 39. For a skeptical discussion of instrumental rationality, *see* Raz, *supra* note 10. Raz takes the position that rationality means response to reasons. An agent's ends may determine which reasons are reasons for that agent, but if the agent's ends lack value, they are not reasons in themselves. Raz's discussion implies that so-called instrumental rationality is best understood as the process of choosing the means for complying with reasons; as such, it is not an independent, self-justifying form of rationality but only one part of the overall process of deliberating about reasons.
32. Shapiro, *Difference*, *supra* note 25, at 39.
33. *Id.* at 47.
34. *Id.* at 47–52.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

agent has no reason to perform the action, it is not rationally available to him. Applying this notion of feasibility to the problem of rule-following, Shapiro draws a distinction between "deliberative reasons," which are reasons to decide upon an act (including the act of committing to a rule), and "implementation reasons," which simply dictate that the agent should do what is necessary to execute a previous decision. Shapiro's key point is that an agent who has committed to follow a rule can no longer *choose* noncompliance because there is no choice left to be made. All decisions to be made were made at T_C, based on a full consideration of "deliberative reasons." Now, at the time of action (T_A), deliberative reasons are no longer in play, and the only reason applicable to the agent is the implementation reason provided by the rule itself. It follows that the only feasible act—the only act the agent can perform for an extant reason—is to comply. This feature of rationality gives agents the power to bind themselves for the future through volitional acts and gives law the power to govern. Shapiro also maintains that rational agents must have this power in order to negotiate the outcome that is best for all present and future selves.[35]

Shapiro's argument for the "causal" power of agents to bind their future selves has the air of wishful thinking: if the power is important enough to realizing one's ends, then it must exist in some literal sense.[36] Yet I do not think that Shapiro intends to say that agents who have committed to a plan actually lack the capacity to defect. Instead, I take him to mean that a *rational* agent will carry through on prior commitments. That is, an agent who holds himself or herself to the standards of rationality will carry through, and an agent who does not carry through is subject to criticism for acting irrationally.[37]

35. The scheme of the argument appears to be this:

(1) An action is not rationally feasible unless it is based on a reason.

(2) An agent's commitment to a plan of action exhausts the force of the deliberative reasons on which the decision to commit was based and replaces them with an implementation reason to follow the plan.

(3) Therefore the deliberative reasons on which the agent's decision to commit was based are no longer available to the agent as reasons for action.

(4) Therefore defection from the plan cannot be based on deliberative reasons pertaining to commitment to the plan.

(5) Therefore defection is not rationally feasible.

36. The term "causal" is Shapiro's. *Id.* at 40. In a discussion of the reasons generated by promises, Joseph Raz make the point that the desirability of a normative power to create reasons for action cannot prove the existence of the power. Joseph Raz, *Is There a Reason to Keep a Promise?* Columbia Public Law Research Paper No.12-320; Oxford Legal Studies Research Paper No. 62 (2012), *available at* http://ssrn.com/abstracts=2162656.

37. Although Shapiro does not qualify the statements he makes about the binding effect of legal plans in LEGALITY, he suggests in earlier work that psychological mechanisms, such as

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

Viewed as an argument about rational deliberation, Shapiro's argument is difficult to defeat on its own terms because it takes off the table anything one might say about deliberation at T_A. At this point, deliberation has ceased, and implementation is the only option. In the discussion that follows, I first point out some potential difficulties that are internal to Shapiro's account. Then, in the final section of the essay, I pose a broader challenge to Shapiro's understanding of rationality.

The most significant problem that arises within Shapiro's account of rational commitment to legal plans is whether and when the agent can escape the effects of commitment. Shapiro suggests at several points that the effect of commitment to a plan on a rational agent is less than absolute. "[C]hoosing a plan," Shapiro says, "does not set it in stone. Reconsideration is rational when, but only when, there is good enough reason for it."[38] What counts as a good-enough reason is not precisely defined, but Shapiro suggests that the reason for reconsideration must be "compelling:" "it would defeat the purpose of having plans if [the agent] were to review their wisdom without an otherwise compelling reason to do so."

An initial question is whether Shapiro really means what he says about reconsideration of plans, at least in the context of legal plans. He evidently does not want to suggest that all laws are absolutely binding on those who accept the basic plan for a system of law. But he equivocates about whether exceptions to the binding effect of laws must be built in to the laws themselves or instead stand as qualifications to the commitment that follows from an agent's acceptance of the basic plan. For example, he states that settlement by means of legal planning:

> should not be taken to mean that the law demands that its dictates be followed *come what may*. Laws, like all plans, are typically defeasible. When compelling reasons exist, the law will normally permit its subjects to reconsider its direction and engage in deliberation on the merits. The catch here is that the law claims the right to determine the conditions of its own defeasibility.[39]

This passage can be read to mean that the defeasibility of commitment is not a feature of rationality but a contingent feature of the law itself. Legal plans may or may not permit agents to reconsider their commitment in exceptional circumstances; when they do not, an agent's acceptance of a legal plan continues to foreclose consideration of any and all deliberative reasons for failing to comply with its terms. Alternatively, the passage can be read to mean either that when an agent commits to follow the basic plan, the agent's former self adds a caveat for problem cases that may arise in the future, or that rationality itself incorporates a caveat for problem cases.

repression of ordinary emotional responses to error, may be needed to make rule-following possible in fact. *See* Shapiro, *Difference*, *supra* note 25, at 52–54.

38. SHAPIRO, LEGALITY, *supra* note 1, at 124.

39. *Id.* at 202. Shapiro refers here to the necessity defense in criminal law.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

The possibility of absolute commitment is not a problem for Alexander and me because we do not believe in commitment.[40] Shapiro does believe in commitment and so must deal with its effects. For one who, like Shapiro, accepts that the prior self can bind the future self, it would seem to follow that the prior self can bind the future self without exception. If Shapiro is willing to bite this bullet, fine. But if not—if he means to say that some sort of escape route from commitment is built into rationality itself—he must spell out in more detail when and why the agent is free to defect.

I assume that the reason for providing an escape route is that it is not rational for the former self to foreclose all consideration by the future self of deliberative reasons for action, no matter how pressing. If this is what Shapiro intends, his first task is to define the type of reason a rational former self would treat as a compelling reason for future defection from a plan. To make the problem concrete, suppose that a group of library patrons wants to maximize their enjoyment of reading and their ability to study, so they appoint a director to make a plan.[41] The director determines that quiet is important to study and enjoyment, that speech within the library is less important, and that if some patrons speak, so will other patrons who would otherwise prefer a norm of silence. For these reasons, the director adopts a rule forbidding speech in general terms: "No talking in the library."

Assuming that Shapiro intends that the "compelling reasons" exception is a built-in qualification on commitment, one way to understand the exception is that patrons have reason to reconsider their commitment to the plan when the plan is not (or has ceased to be) instrumentally justified as a rule. If circumstances have changed so that the plan no longer satisfies NERP, those who have committed to follow it are released from rational obligation. Perhaps the plan for the library worked well for some time, but it is now 2020, and the library no longer contains books (it contains rows of computers that display digital material and are equipped with noise-blocking headphones). As a result, the overall benefits of consistent application of the rule no longer outweigh the benefits of speech, and patrons are no longer bound to comply. This version of the compelling-reason exception, however, is quite narrow, particularly if the NERP test is applied not at the level of specific laws but at the level of the master plan for the legal system.[42]

40. More accurately, we do not believe in commitment *per se*. Commitment can have moral implications if it appears that others have relied on the agent's commitment. But in this case reliance, not commitment itself, counts as a reason to carry through on the commitment. *See infra* notes 49–51 and accompanying text.

41. In previous work, Alexander and I use a similar library example to illustrate points about interpretation. Alexander & Sherwin, Rule of Rules, *supra* note 8, at 114–116.

42. A further problem is that the overall justification for the rule is probably best determined by officials with authority to overrule the rule rather than by individual agents considering how to respond to particular cases. Agents may be relatively poorly informed and distracted by the facts of the cases they face, and overruling has coordination advantages.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

Shapiro seems to have in mind a more general exception that would apply to justified as well as unjustified rules.[43]

Another way to interpret the compelling-reason exception is as a safety valve for extreme cases in which compliance could have disastrous consequences. A library patron, noticing that the person seated next to him has suffered a heart attack, decides to call 911. This is certainly a worthy reason to disregard the no-talking rule, and perhaps the "disaster" category can be cabined in a reasonable way. It does not seem necessary, however, to invoke an exception to the patron's commitment to follow the rule in order to avoid truly disastrous results. The same result can be reached by interpretation, at least under an interpretive theory in which the meaning of a rule turns on the rule-maker's intent.[44] One can easily say, without engaging in creative or purposive interpretation, that a library director who issues a rule "No talking in the library," with the general aim of preserving a quiet environment for study, does not intend to forbid lifesaving phone calls. Again, however, I doubt that this understanding of compelling reasons for defection would satisfy Shapiro: he seems to anticipate that the former self would give the later self a somewhat longer leash than life-threatening emergency.

The more difficult case, and the case Shapiro probably has in mind when he speaks of compelling reasons to reconsider, is one in which the rule is well justified as a rule, and no dire emergency has arisen, but the agent's reasons to defect strongly outweigh reasons to honor a commitment to follow the plan. A patron learns while browsing the news on a cell phone that the president of the United States has just been assassinated. This is not a personal emergency, and announcing the event will create a disturbance, but the patron may judge that most others in the room would wish to know. So she speaks. Arguably, at least, this patron has a compelling reason both to reevaluate her reasons for accepting the rule and to violate the no-talking rule by announcing what has occurred. I am not sure whether Shapiro would accept this as a case for reconsideration of commitment,

43. The exception for compelling reasons might also be understood along lines suggested by Joseph Raz. Raz depicts legal rules as sources of "exclusionary reasons" for action—second-order reasons that exclude from deliberation the set of first-order reasons on which the rules depend. He indicates, however, that the exclusionary reason provided by an authoritative rule excludes only those first-order rules on which the rule-making authority relied in issuing the rule. In effect, reasons on which the authority relied are merged into the rule, but other reasons may still be in play. By analogy, Shapiro's "compelling reasons" might be those that the lawmaker did not consult in deciding to issue the rule. This, too, seems too narrow an exception for Shapiro's purposes; and in any event, how it would apply in practice is something of a mystery. *See* JOSEPH RAZ, THE AUTHORITY OF LAW (1979), at 16–19, 22–23, 30–33; JOSEPH RAZ, THE MORALITY OF FREEDOM (1986), at 57–62; Joseph Raz, *Authority & Justification*, 14 PHIL. & PUB. AFF. 3 (1985) (giving the example of a rule based solely on economic considerations).

44. This is the position Alexander and I advocate. *See* note 26, *supra*; ALEXANDER & SHERWIN, RULE OF RULES, *supra* note 8, at 96–122. What we mean by intent is not the purpose for which the lawmaker issued the rule but the meaning the lawmaker intended the words of the rule to have. What meaning the lawmaker intended is, of course, a complicated question, but in the case of emergencies, the interpretive problem is not too difficult.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

but he might: something unusual has happened that significantly alters the balance of the reasons on which the rule is based. The patron is not simply recalculating but responding to an unforeseen contrary reason of considerable strength.[45]

Another preliminary issue is whether the existence of a compelling reason to defect from commitment is an objective question or a question to be answered from the internal perspective of the agent. I am already assuming that Shapiro is applying a standard of rational deliberation to which agents will attempt to conform and by which they can be criticized if they fail to conform. Within this framework, it seems appropriate that the exception for compelling reasons should at least track the evidence available to the agent. Whether it should also allow for idiosyncrasies in the agent's evaluation of evidence is not so clear.[46]

I believe the best reading of Shapiro's comments about exceptions to commitment is that agents are permitted to defect when they believe or reasonably believe that the reasons to defect strongly outweigh the reasons to obey (including the value of binding commitment). Permission to defect derives from the standards of instrumental rationality, which dictate that the self who makes a commitment can impose a substantial obligation of compliance but not an absolute obligation of compliance on future selves. Given these assumptions, the most serious problem Shapiro faces is that an exception that allows agents to defect from commitment in circumstances not authorized by the rule itself is hard to reconcile with Shapiro's explanation of the causal force of commitment. In particular, an exception of this kind appears to undermine the distinction Shapiro draws between "deliberative" reasons and "implementation" reasons. Shapiro asserts that at the point of application of a plan (T_A), there is no decision open to the agent; therefore deliberative reasons are no longer in play. Yet, if the agent believes that there may be reasons to defect in a particular case, it seems that the agent must engage with deliberative reasons in order to determine whether those reasons are strong enough to support a release from commitment.

This is true under any of the various definitions of "compelling" reasons discussed above. Reconsideration of the rule or plan under NERP requires the agent to review the set of deliberative reasons supporting the plan as well as the errors the plan may cause in application. Reconsideration of commitment in case of emergency requires the agent at least to assess the deliberative reasons for avoiding disaster and probably also to compare them to the deliberative reasons supporting adherence to the plan. Reconsideration of commitment when some new application of a justified rule or plan creates a strong reason to depart from the plan requires the agent to

45. This roughly tracks the approach Frederick Schauer defends—an approach Schauer refers to as "presumptive positivism." SCHAUER, *supra* note 27, at 202–205.

46. *See* Greco, *supra* note 10 (suggesting that the appropriate standard of "epistemic probability" for purposes of rational choice should reflect the evidence actually available to the agent but should exclude idiosyncratic conclusions the agent may draw from that evidence).

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

consider both the deliberative reasons underlying the rules and the deliberative reasons for disobedience in order to determine whether the reasons for disobedience are comparatively strong. Thus, unless Shapiro takes the strict view that only the law itself can determine when reconsideration is allowed, he must concede that some sort of decision involving deliberation goes on at the point of applying a plan.

Further, even if Shapiro limits defection to cases in which authorization to reconsider for compelling reasons is written into the plan, there are questions about how a vague exception of this kind can remain stable. Suppose the stated exception tracks the third possibility described above: agents have a compelling reason to depart from the plan if the circumstances of particular case generate a strong reason to do so. The difficulty is that agents will often believe that they have strong reason to defect, and they will sometimes be wrong. Further, even a correct application of the exception can rapidly undermine the coordination effects of a rule. For example, consider the rule "No talking in the library," now amended to read "No talking in the library without a compelling reason to speak." Patron A sees what appears to be a heart attack, determines that he has a compelling reason to speak, and phones 911. Patron B receives a call from home, where a babysitter is watching her child, and considers whether she has a compelling reason to speak. B's reason to speak may be less compelling than A's reason to speak, but at this point the threshold is lower because there is already some noise, or a recent history of noise, in the library. If the pattern continues, the level of cell phone chatter may reach a point at which almost any reason to answer a call looks compelling when compared to potential harm to the rule's value in preserving an environment conducive to study.

IV. RATIONALITY PLUS

Suppose the problems I have raised so far can be ironed out. There remains a basic divide between Shapiro's understanding of rationality and the understanding of rationality that leads Alexander and me to question the ability of rules to constrain conduct. Shapiro views constraints put in place by an agent's former self as limits on the reasons for action available to the agent's future selves, while Alexander and I believe that agents can, and as a matter of rationally should, respond to all reasons that apply to them at the time of action.[47] In our view, the final decision to act is always a decision, and there is no mechanism by which prior intentions can curtail deliberation about

47. I am setting aside questions about how reasons for action can actually cause actors to act. *See*, generally, Jay R. Wallace, *Practical Reason*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY 8–10 (Edward N. Zalta ed., 2009), *available at* http://plato.stanford.edu/archives/sum2009/entries/practical-reason/.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

that decision. I refer to Shapiro's view as the model of constraint,[48] and to Alexander's and my view as the model of full deliberation.

Shapiro's model of constraint has the advantage of enabling human agents to adhere without the aid of psychological distortions to plans that will advance their ends more effectively than unconstrained decision-making at the time of action. If true, this is good news: reason itself can maximize success in pursuit of one's ends. I argue, however, that the model of full deliberation gives a better picture of rationality, because it acknowledges not only standards of instrumental rationality but also standards governing the broader process of deliberative decision-making. In particular, full deliberation takes account of epistemic responsibility, a consideration that Shapiro does not address.

I begin by noting a move that Shapiro does not make. It can be tempting to think that commitments, promises, and the like create additional reasons to honor the commitment or promise over and above whatever initial reasons the agent had to do the planned or promised act.[49] In my view, this is a mistake.[50] Suppose, for example, that the balance of reasons narrowly favors that I work tonight rather than watch a movie. I then make a plan to rent a DVD and commit to following through on the plan. It is difficult to see how my commitment can add anything to my reasons for watching a movie, which continue to be outweighed by my reasons to work.[51]

Shapiro's analysis is consistent with this observation. Shapiro claims that commitment generates an "implementation" reason to follow through on plans, but, as I interpret his argument, the implementation reason does not add to the balance of reasons favoring conformity to a particular plan. Instead, the new implementation reason displaces the set of deliberative reasons that previously favored the plan. If these reasons were outweighed initially by the reasons against following the plan, the implementation reason that arises from commitment would also be outweighed. Nothing has changed, and it is not rational for the agent to follow the plan.[52]

48. Shapiro uses this term in his extended essay on the subject. *See* Shapiro, *Difference*, *supra* note 25, at 39.

49. For discussion of content-independent reasons for action, *see* H.L.A. HART, ESSAYS ON BENTHAM (1982); Raz, *Is There a Reason*, *supra* note 36.

50. For an argument against the existence of content-independents reasons, *see* P. Marwick, *Law and Content-Independent Reasons*, 20 OXFORD J. LEGAL STUD. 579 (2000). Raz makes the point that a reason with no "strength-determining factors" cannot in itself be a reason for action even if the existence of the reason would have significant benefits. Raz, *Is There a Reason*, *supra* note 36, at 14, 16. In the case of promising, Raz locates strength determinants in the value of the normative power to make particular binding commitments; *id.* at 25–27.

51. If the commitment takes form of a promise to a third person, the promise may lead the promisee to take action in reliance or may have other consequences that create reasons to perform. Reasons of this kind, however, depend on the effects of the promise, not on the promise itself. *See* Raz, *supra* note 36, at 7.

52. This interpretation avoids any questions about whether standards of rationality should distinguish between good and bad reasons for action. A plan based on bad deliberative reasons generates a bad implementation reason, which is no reason at all.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

Shapiro's mistake lies instead in his failure to consider the rationality of beliefs. An agent facing the question whether to follow or defect from a plan or rule must form a belief about which course of action the agent ought to pursue. Perhaps discussion of belief formation simply shifts the original debate from the level of reasons for action to the level of grounds for belief. Thus Shapiro might say that an agent who has committed to follow a rule will believe that he or she ought to follow the rule given that deliberative reasons to the contrary are not available. Alexander and I would say that the agent should form whatever belief the full array of reasons recommends. We remain at an impasse. Arguably, however, focusing on the agent's beliefs changes the picture of rule-following in two ways.

First, standards of epistemic responsibility may not correspond to standards of instrumental rationality. According to traditional accounts of epistemic responsibility, beliefs are justified when they are supported by the evidence available to the agent.[53] The traditional view has met with some dissent, but most would agree that the immediate purpose of epistemic standards is not to enable the agent to advance his or her ends but to enable the agent to acquire true beliefs and avoid false beliefs. Consequently, standards for justified belief will be different from and may not always be consistent with standards for instrumental rationality.

Along these lines, Thomas Kelly makes a strong argument that the fact that an agent's holding a certain belief will serve the agent's ends cannot justify the agent in adopting the belief. Kelly assumes that it is possible to affect one's own beliefs, for example by seeking out evidence that confirms a particular belief. Yet he argues, plausibly I think, that a belief acquired in this way is based not on the instrumental value of the belief but on the new evidence the agent acquires in support of the belief.[54] Kelly gives content to the "based-on" relation by asking whether the agent would continue to hold the belief if its instrumental value decreased. Thus, if Pascal thinks it may be useful to believe in God, he may seek out evidence that supports belief in God and eventually come to believe in God. But if he later changes his mind about the advantages of believing in God, he will not simply stop believing in God; to stop believing, he will need to seek contrary evidence that actually alters his belief. Accordingly, Pascal's belief in God is not *based on* the benefits of holding the belief. Kelly concludes that "[w]ith respect to beliefs, rationality just is epistemic rationality"[55]—which he takes to mean responsiveness to evidence rather than responsiveness to consequences.[56]

Assuming that standards for justified belief are different from standards for instrumental rationality, as Kelly's argument indicates, they are also

53. *See*, generally, Matthias Steup, *Epistemology*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY 3–4, 8–9 (Edward N. Zalta ed., 2013), *available at* http://plato.stanford.edu/archives/win2013/entries/epistemology/.

54. Thomas Kelly, *The Rationality of Belief and Some Other Propositional Attitudes*, 110 PHIL. STUD. 163 (2002), at 171–177.

55. *Id.* at 171.

56. *Id.* at 165.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

naturally prior to standards for instrumental rationality. If an agent fails to conform to standards of belief formation that are reasonably calculated to promote true belief, deliberation about the relations of means to ends is unlikely to result in practical success. Thus, to put practical instrumental reasoning on a firm footing, the agent must first form relevant beliefs in an appropriate way. For present purposes, I am assuming, in keeping with traditional views of epistemic responsibility, that this means the agent's beliefs are supported by the agent's evidence.

The second way in which focusing on epistemic responsibility alters the debate over rule-following is that it illuminates the distinction between general and particular responses to plans or rules. When an agent deliberates about whether to commit to follow a rule, the agent considers the rule in the abstract as it will apply to the full set of cases that may arise over time. When the question is whether to follow the rule, the agent considers only one application of the rule. At this point, the array of evidence available to the agent as well as the relevance of individual items of evidence is different from what it was at the earlier time, when the agent committed to follow the rule.

To see this, suppose again that the director of the library where agent A is a frequent patron has issued a rule, R ("No talking in the library"). Implicit in rule R is the further instruction that R is to be applied in all cases covered by its terms. At time T_C, when A considers whether to accept this rule, the only relevant evidence—that is, the only evidence that could make a difference to A's decision—is evidence that bears on R's long-term effects. For example, suppose A knows that A and other library patrons prefer quiet to speech within the library; that library patrons are unable to coordinate their behavior spontaneously; that A's reasoning and the reasoning of other patrons are fallible; and that the time A can devote to future deliberation is limited. A also knows that R is imperfect: in some future cases, speech will be more important than both silence and the settlement benefits that may be lost through defection. The errors R may generate, however, matter only insofar as they affect R's satisfaction of NERP. If NERP is satisfied, A knows that despite some errors, A will do better over the long run by following R consistently than by following her own judgment. Based on this evidence, A concludes that she can best advance her ends by accepting rule R and committing to follow it in all cases that come within its terms. This conclusion, it is fair to assume, is well supported by A's evidence.

At time T_A, A receives a phone call that appears to be from her babysitter. A must (quickly) form a belief about whether she ought to answer the call immediately or follow the rule and call back from another location. A now has new evidence: she knows that she has received a phone call, believes the phone call may be urgent, and believes she can handle it quickly. After glancing around the room, she perceives that there is only one other patron present, which indicates that a quick violation of R would do minimal

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

damage to the coordination powers of the rule. So A concludes that following R would be an error.

In a sense, A's evidence is not new—or at least not unanticipated—because A has known from the outset that following R in every case would cause some errors. A also knows at T_A what she knew at T_C about the long-term benefits of following R. But the change in her vantage point has altered the weight of this evidence. The long-term benefits of R are pertinent, but only in comparison to the particular evidence of error A now has (a call from the babysitter, etc.) and only to the extent that the particular violation A is considering will undermine the settlement value of R. In other words, NERP is not relevant to the question whether, in *this case*, the error of following R exceeds the error of violating R. A's commitment to follow R continues to be instrumentally rational, but if A concludes that the particular evidence before her favors a violation of R, she cannot believe, consistent with standards of epistemic responsibility, that she ought to follow R.

At this point, Shapiro is likely to say that A's beliefs, like A's actions, should conform to the model of constraint: because A has committed to follow R, R now dictates what A should believe about what she ought to do. Following R is instrumentally rational; holding the belief that she ought to follow R will enable A to follow R; therefore holding the belief that she ought to follow R will advance A's ends; therefore A should believe that she ought to follow R. But this takes us back to Pascal's belief in God: from the fact that A can advance her ends over time by believing that she ought to comply with R, it does not necessarily follow that A is epistemically *justified* in believing that she ought to comply with R. As Kelly shows, justification for belief depends on evidence, not consequences.

Accordingly, I do not think that Shapiro can avoid the problem of epistemic responsibility at T_A. When an agent faces a particular application of a plan or rule, the difference in the agent's vantage point alters both the evidence available to the agent and the importance of evidence about the long-term benefits of the plan or rule. In the library example, it is hard to see how A, as an epistemically responsible agent, can ignore the evidence that following R may lead her to err. If, taking account of this evidence, A comes to believe that she ought not follow R in this case, her belief may be wrong but it will be justified. If, instead, she forms the belief that she ought to follow R in this case because she has committed to follow R, and acts on her belief, her act of rule-following may be instrumentally rational but the belief on which it was based will not be justified.

My own sense, at least, is that the second of these—lack of justified belief—represents a more serious flaw in the overall process of A's reasoning. Instrumental reasoning that rests on irresponsible beliefs should not count as rational, no matter how useful or even essential it is to successful management of social life. Consequently, I adhere to the position Alexander and I have taken—that although it may be eminently rational to endorse a rule as a rule to be followed in all cases, it is not always rational to follow it in fact.

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press

If I am right that epistemic standards prevent agents such as A from following rules, this is not just a complication for Shapiro's model of constraint but a challenge to the core idea of rational commitment. If A deliberates carefully at time T_C, A will understand that she cannot form a justified belief that she ought to follow R when the evidence available to her in a particular case indicates that following R would be an error. Accordingly, she will understand that in some future cases she will not be able to follow R at T_A, despite her commitment at T_C.[57] This makes the commitment itself futile and therefore irrational by Shapiro's own standards of instrumental rationality.

V. CONCLUSION

I agree with much of what Shapiro says in his excellent book—this much should be clear from a comparison of our treatments of the subject of law and settlement. Nevertheless, I argue that Shapiro's constraint model of rationality, which supports his faith in the capacity of legal plans to govern action, is incomplete. The result is a highly convenient but ultimately unworkable view of the effect of legal plans on agents who accept them. Thus I continue to believe with Alexander that the effectiveness of law as a means of settlement depends on a fundamentally irrational willingness of most agents to follow the law.

57. This is similar to Gregory Kavka's toxin puzzle, in which a billionaire offers to pay you $1 million if you intend tonight to drink a vial of toxin tomorrow, adding that you do not have to drink the toxin in order to win the money. Can you form the necessary intention? Gregory Kavka, *The Toxin Puzzle*, 43 ANALYSIS 33 (1983). Bratman, defending a theory of intentions similar to Shapiro's model of constraint, deflects this example by arguing roughly that it is rational to induce oneself to intend tonight to drink tomorrow by some artificial means (such as self-hypnosis). But the ability of an agent to form an intention and resist reconsideration is limited to intentions formed deliberatively; therefore the agent remains free to reconsider an intention formed by nondeliberative means. Therefore Bratman does not have to concede that is rational to drink the toxin. *See* BRATMAN, *supra* note 1, at 101–106. This argument, however, elides the question whether it is possible to form an intention to drink deliberatively (without the aid of hypnosis).

https://doi.org/10.1017/S1352325213000219 Published online by Cambridge University Press