Legal Theory, **12** (2006), 315–345. Printed in the United States of America
Published by Cambridge University Press 0361-6843/06 $12.00 + 00

INSTRUMENTAL RULES AND MOTIVATION*

Antony Hatzistavrou
University of Cyprus and Clare Hall, Cambridge University

In this paper I address the issue of the normativity of instrumental rules (for example, legal rules). On the one hand, I criticize Scott Shapiro's "constraint model" of instrumental rule-following according to which instrumental rules have motivational clout qua rules: the agent conforms to them *simply because* they are rules. On the other, I argue for a purely epistemic account of instrumental rule-following. According to this account, instrumental rules inform the agent which action she is required to perform but do not get her to act. Rather the agent is motivationally guided by the reasons for which she adopted the instrumental rules.

In a series of recent articles[1] Scott Shapiro tries to capture the normativity of instrumental rules[2] in terms of their effect on the motivation of the rule-follower. He develops a particular account of the motivation of the instrumental rule-follower, the constraint model. According to this model, the main function of rules is the exercise of volitional control over our future selves by rendering nonconformity infeasible. An essential tenet of the constraint model is the practical-difference thesis. According to this thesis, when the agent conforms to the rule she believes that were she not

*Earlier versions of this paper were presented at a meeting of the Stirling Political Theory Group and at an Open Session of the 2006 Joint Meeting of the Mind and the Aristotelian Society at the University of Southampton. I am grateful to the participants of these events (especially Maria Alvarez, Antony Duff, Kathlyn Farkas, and Kent Hurtig) for comments and suggestions. I am also grateful to Matthew Kramer and two anonymous referees of LEGAL THEORY for their written comments on an earlier draft. I owe a special debt to Scott Shapiro for a long discussion on his theory and a set of written comments on an earlier draft.

1. *See* Scott J. Shapiro, *On Hart's Way Out*, 4 LEGAL THEORY 469 (1998); Shapiro, *The Difference That the Rules Make, in* ANALYZING LAW: NEW ESSAYS IN LEGAL THEORY (Brian Bix ed., 1998); Shapiro *Law, Morality and the Guidance of Conduct*, 6 LEGAL THEORY 127 (2000).

2. Shapiro's arguments concern exclusively instrumental rules (*see* Shapiro, *The Difference That The Rules Make*, *supra* note 1, at 36). A gloss on "instrumental": I do not take it to range only over relations of means to an end but also over constitutive relations and relations of instantiation. For example, observing articles of the constitution is constitutive of respecting the sovereignty of the state; and following promise-keeping rules is an instance of having a promise-keeping attitude. I treat "instrumental" as roughly equivalent to "conducive" in the way suggested by ALFRED MELE, MOTIVATION AND AGENCY 21 (2003). For the sake of simplicity I speak in the paper only of "means to end" relations, but my arguments should be understood to hold for the other two relations as well.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

so motivated by the rule, she might have failed to conform to the rule.[3] Further, Shapiro provides an original interpretation of Hart's account of rules as peremptory and content-independent reasons for action along the lines of the constraint model. On this interpretation, rules as peremptory and content-independent reasons for action volitionally constrain the agent to conform to the rule and make her believe that nonconformity is not an option for her.

In this paper, I advance two arguments against the constraint model. The first is that the practical-difference thesis fails to capture an essential feature of the relation between motivation and behavior that I call the "motivational determination of action by reasons." A reason motivationally determines action not simply if it inclines the agent toward the action but if it is the reason why of the action. As I explain in this paper, the practical-difference thesis fails to distinguish an inclinatory reason from the reason why of an action in cases in which the agent has two reasons for performing an action but does it for one reason rather than the other.

The second is that the constraint model does not provide a convincing account of how rules volitionally control the agent. I argue that rules as peremptory and content-independent reasons for action are motivationally inert. They can provide only epistemic guidance: they inform the agent which action she is required to perform. But they do not provide motivational guidance. What gets the agent to conform to the rule is rather the desire for the ends to which she believes the rule to be instrumental.[4] This leads to the positive conclusion of my paper: in "standard" cases of instrumental rule-following[5] the reasons for adopting the rule motivationally determine conformity with the rule.

Two points of clarification are needed concerning the scope of my argument. First, I am exclusively concerned with Shapiro's account of the normativity of instrumental rules. Shapiro builds on his account of rule-following a defense of exclusive legal positivism. In this paper I neither discuss Shapiro's defense of exclusive legal positivism nor explore the implications of my criticisms of Shapiro's argument for the debate between exclusive and inclusive legal positivism.[6] The constraint model is an original theory about the function of rules in practical deliberations that deserves

3. In the remainder of the paper, "rule" and "rule-following" stand for "instrumental rule" and "instrumental rule-following," respectively.

4. Though I imply here a causalist account of reasons for action, as I will indicate later, I intend my arguments to be neutral as to whether beliefs or only desires are motivational states and as to the truth of causalist or teleological accounts of reasons for action. For convenience of exposition, in the third part of the paper I mainly talk only of desires as motivational states (assuming thus a causalist account).

5. I explain what I mean by "standard" cases in Section III; cf. note 50.

6. For criticisms of Shapiro's defense of exclusive legal positivism, *see* Kenneth Himma, *H. L. A. Hart and the Practical Difference Thesis*, 6 LEGAL THEORY 1 (2000); Will Waluchow, *Authority and the Practical Difference*, 6 LEGAL THEORY 45 (2000); and Matthew Kramer, *How Moral Principles Can Enter into the Law*, 6 LEGAL THEORY 83 (2000).

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

to be studied in its own right. For it provides an answer to the question of how law affects the behavior of its subjects. So Shapiro's theory of practical deliberation is of importance for jurisprudence independently of its use as an argument for exclusive legal positivism.

Second, Shapiro blends the presentation of his theory of instrumental rule-following with a reconstruction of Hart's internal point of view and Hart's account of rules as content-independent and peremptory reasons for action. In this paper I do not address the issue of the exegetical value of Shapiro's interpretation.[7] That is, I do not examine whether Hart actually thinks that the internal point of view or the peremptory and content-independent reasons have the kind of motivational clout ascribed to them by Shapiro's interpretation. I am concerned only with whether the constraint model correctly captures the essence of instrumental rule-following and not with whether it correctly captures the gist of Hart's account. Consequently, when in this paper I speak about the internal point of view or the peremptoriness and the content-independence of rules, I refer only to how Shapiro understands them. Whether Hart actually intends them the way Shapiro understands them is immaterial to my argument.

I. MOTIVATIONAL GUIDANCE AND THE CONSTRAINT MODEL

Shapiro distinguishes between epistemic and motivational guidance of legal rules in the context of his discussion of Hart's internal point of view. On the one hand, "A person is epistemically guided by a legal rule when the person learns of his legal obligations from the rule provided by those in authority and conforms to the rule." On the other hand, a person is motivationally guided by a legal rule if, in addition, that person's conformity with the rule "is motivated by the fact that the rule regulates the conduct in question." Shapiro contrasts epistemic with motivational guidance as follows:

> [in the case of epistemic guidance] it is not necessary that the agent be motivated to follow the rule *because* of the rule. A rule can epistemically guide conduct even though compliance is motivated by the threat of sanctions, just in case the person learns from the rule what [s]he must do in order to avoid being punished.[8]

To clarify the distinction Shapiro draws between epistemic and motivational guidance, I introduce a third notion, that of mere epistemic guidance. If X does A (which counts as conformity to rule R) in circumstances C and

7. I explore this issue in Hatzistavrou, *An Epistemic Reading of the Internal Point of View* (forthcoming). I argue that, contra Shapiro, Hart's internal point of view should be equated with epistemic rather than motivational guidance.

8. *See* Shapiro, *On Hart's Way Out*, *supra* note 1, at 490 emphasis in original).

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

believes that R applies in C and requires that X does A, then X is epistemically guided by R (irrespective of what motivated X's doing A). If X does A (conforms to R) in C and believes that R applies in C and requires that X does A and X is motivated to do A because of R, then X is not only epistemically guided by R but also motivationally guided by R. If X does A (conforms to R) in C and believes that R applies in C and requires that X does A and X is motivated to do A *not because of* R but for some other reason, then X is merely epistemically guided by R. For example, if I abstain from jaywalking because there is a legal rule against it, I am motivationally (and thus also epistemically) guided by the relevant legal rule. If, nevertheless, though I know that the legal rule forbidding jaywalking applies to my case, I abstain from jaywalking not because of the legal rule but in order to avoid possible sanctions, I am merely epistemically guided by the legal rule. In the latter case I learn from the legal rule what my obligations are (and thus what I should do in order to avoid punishment) but I am not motivated to conform because of the rule but because of my desire to avoid punishment.[9]

On Shapiro's interpretation, Hart believes that in order for a legal system to exist, the legal officials must be motivationally guided by law. To be motivationally guided by law is to adopt the internal point of view. But it is not necessary for the existence of a legal system that the ordinary citizens are motivationally guided by legal rules. It suffices that they are (even merely) epistemically guided by law.[10] Ordinary citizens need only know what the law requires and obey.

The equation of motivational guidance with the possession of the internal point of view indicates that motivational guidance rather than mere epistemic guidance is fundamental for Shapiro's account of the normativity of rules. As the internal point of view, or more precisely the notion of the internal aspect of rules, was fundamental for Hart's account of rule-following, so is motivational guidance for Shapiro's account of rule-following.[11] For

9. I do not accept an anonymous referee's suggestion that, for Shapiro, if I conform to the rule that forbids jaywalking in order to avoid possible sanctions, I am motivationally guided by that rule. First, it goes against the gist of Shapiro's juxtaposition of epistemic guidance with the fear of sanctions in his statement that "a rule can epistemically guide conduct even though compliance is motivated by the threat of sanctions, just in case the person learns from the rule what [s]he must do in order to avoid being punished." (For a similar juxtaposition, *see* the first paragraph in *id.* at 492.) I think what Shapiro means is that the person who is motivated by the threat of sanctions is not motivationally guided by the rule but is (merely) epistemically guided by it. The confusion is created by the fact that in these passages Shapiro uses the concept of mere epistemic guidance but does not explicitly mark the use of this concept. Second, the suggestion goes against Shapiro's claim that norm-guidance (which, as I explain in note 11, must be considered equivalent to motivational guidance) differs from a power relation (*id.* at 473). If being motivationally guided by a rule included being motivated out of fear for possible sanctions, then there would be no difference between adopting the internal point of view and being a subject of the Austinian sovereign.

10. *Id.* at 491–492.

11. Shapiro is not explicit about how epistemic and motivational guidance are related to his notion of norm-guidance with which he introduces his discussion of Hartian jurisprudence in *On Hart's Way Out.* Norm-guidance is contrasted to norm-governance. A norm-governed

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

this reason I direct the bulk of my criticisms to the workings of Shapiro's account of motivational guidance.

In a later article Shapiro clarifies that a complete account of the distinction between epistemic and motivational guidance should take notice of the fact that agents appeal to rules under several descriptions.[12] For example, an agent might appeal to a rule prohibiting murder as a moral and not as a legal rule. And one may be motivationally guided by that rule qua *moral* rule but not qua *legal* rule. That is, one may abstain from murder because it is morally prohibited and not because it is legally prohibited. Similarly, one might learn from it of one's moral and not of one's legal obligations. In this case one is neither motivationally nor merely epistemically guided by the rule against murder *as a legal rule*.

According to Shapiro, two conditions are satisfied when the agent is motivationally guided by legal rules qua legal rules. The first is that the agent acts on the belief that the rule is a legitimate standard of conduct.[13] That is, the agent believes that she ought to conform to the rule. The second is that the adoption of the rule makes a practical difference to the deliberations of the agent. The agent believes that had she not adopted the rule, she might not have acted in accordance with the content of the rule.[14]

behavior is a pattern of behavior regulated by an actual norm (for example, if there is a norm prohibiting jaywalking, then crossing the street is a norm-governed behavior). It is irrelevant whether anyone conforms to the norm for the relevant pattern of behavior to be norm-governed. Even if no one conforms to the no-jaywalking rule, crossing the street remains a norm-governed behavior. By contrast, a pattern of behavior is norm-guided only if one conforms to the norm because the norm regulates that behavior. So only if I abstain from jaywalking because there is a rule prohibiting it is my crossing the street norm-guided. For Shapiro, Hart solves the puzzle concerning the origin of the legal authority of law (what makes legal rules law) by taking all norm-governed behavior (i.e., all legal rules) to depend on norm-guided behavior (i.e., the guidance of legal conduct by the rule of recognition) (*id.* at 472–475). This strongly suggests that for Shapiro norm-guidance is motivational guidance and is not equated with mere epistemic guidance.

12. Shapiro, *Law, Morality*, *supra* note 1, at 132. There is no inconsistency between the account of motivational guidance offered by Shapiro in this article and the account offered in his other two articles cited in note 1, as an anonymous referee suggests. In particular Shapiro is not using "motivation" in *Law, Morality and the Guidance of Conduct* in a sense that comprises both epistemic and motivational guidance. It is not the case for Shapiro that anything that guides moves. Rather only what motivationally guides motivates. What epistemically guides may or may not motivate (as it happens in cases of mere epistemic guidance). This becomes explicit with Shapiro's remark about the epistemic guidance of the law: "the guidance function of the law is served regardless of whether people conform out of a sense of moral obligation, tradition, habit or fear" (*id.* at 147; that Shapiro refers here to epistemic guidance is clear by the context). The confusion is, I think, again created by the fact that Shapiro does not explicitly draw a distinction between epistemic guidance (which may be motivational as well) and mere epistemic guidance (which is not).

13. "To be motivated to conform to a legal rule by the rule itself is to believe that the rule is a legitimate standard of conduct and to act on belief"; Shapiro, *On Hart's Way Out*, *supra* note 1 at 490; cf. n. 26.

14. "[F]or rules to guide conduct, they must be capable of making 'practical differences,' i.e. they must be capable of motivating agents to act differently from how they might have without their guidance"; *id.* at 493.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

To these two conditions we may add a (trivial) third: the agent believes that the rule applies in the circumstances.

The practical-difference thesis is related to the main function of rules in practical deliberation. According to Shapiro, the whole point of having rules is to *constrain* our future selves in following particular courses of action. By adopting a rule, our present self makes certain alternatives infeasible to our future self. This means that my future self believes that she has no choice but to do what my present self, who adopted the rule, decided. If my future self would do the required action without being constrained by the rule, there is no point for me to adopt the rule.

Shapiro contrasts the constraint model with the decision model. On the latter, rules secure that the agent will always *prefer* conformity to nonconformity. The adoption of the rule affects the preferences of the future self. Shapiro explains this point by reference to Raz's notion of exclusionary reasons.[15] The rule excludes certain considerations for nonconformity from entering the balance of reasons and in this indirect way affects the balance. The future self will have to make a decision on what to do, and the adoption of the rule secures that the balance of reasons is always in favor of the action the rule prescribes. So, on the decision model, the future self is considered to have a choice as to whether to conform to the rule. The rule affects the agent's judgement, thus securing that she will choose to conform. By contrast, on the constraint model, the future self does not have a choice (i.e., nonconformity is not an option). The rule affects the volitional control of the present over the future self.[16]

For example, on the decision model, my self-adopted rule to abstain from alcohol on medical grounds functions in the following way in my practical deliberations. When my future self will be at a bar ready to order a drink, she will believe that she has a choice about whether to order alcohol or not. The rule will affect her preferences, securing (e.g., by excluding some conflicting considerations) that the balance of reasons is tipped in favor of ordering a nonalcoholic beverage. By contrast, on the constraint model, my future self will not believe that ordering an alcoholic drink is an option for her. When she recognizes that the rule applies, she will be disposed to conform to it. The rule does not affect her preferences, since she does not consider the situation as one in which she has a choice between conformity and nonconformity. The rule controls the future self's will so that she does what the previous self who adopted the rule decided.

One of the reasons for which Shapiro finds the constraint model superior to the decision model is that it can account for cases of rule-following in which the balance of reasons is insensitive to the adoption of the rule. If we modify our original example by adding that I am a recovering alcoholic, then

15. Again I do not examine the exegetical value of Shapiro's interpretation of Raz's account of exclusionary reasons.

16. For the contrast between the decision model and the constraint model, *see* Shapiro, *The Difference that Rules Make*, *supra* note 1, at 33–38 and 47–54.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

the adoption of the rule to abstain from alcohol would be pointless on the decision model. The balance of reasons will always be in favor of abstaining from alcohol, so the adoption of a relevant rule (as an exclusionary reason) would not make a difference.[17] By contrast, on the constraint model one can understand the rationality of my self-adopted rule. The adoption of a rule makes a difference because it fosters my volitional control over my action. Even though I will always prefer abstention from alcohol, since one does not always act on one's best preference, it is rational for me to adopt a rule and not to rely each time on the balance of reasons. The rule fosters my volitional control over myself.

So on the constraint model rules make a practical difference in the motivation of the agent because they render nonconformity infeasible. Shapiro defines the central thesis of the constraint model as follows:

> **Feasibility Thesis.** An agent is being instrumentally guided by rule R if and only if
>
> 1) the agent's behaviour conforms to R;
> 2) recognition that R applies constrains the agent to conform to R, making non conformity infeasible; and
> 3) the agent believes that absent the motivation of R, [s]he might not conform to R.[18]

The scope of the belief in condition 3 is ambiguous;[19] 3 could be taken as equivalent to either:

> 3*) the agent believes that, absent the motivation of R, she might not conform to R in the specific case C_1 she is currently facing;
> or
> 3**) the agent believes that there is at least one possible case C_p (though not necessarily the one she is currently facing) in which, absent the motivation of R, she might not conform to R.[20]

17. *Id.* at 35. It should be noted that Shapiro does not speak here of the "perceived" balance of reasons and does not claim that my recovering alcoholic self cannot make a mistake about the balance of reasons. He refers to the rationality of adopting the rule. His point seems to be that given my condition as a recovering alcoholic, even if I include in my deliberations the reasons for drinking, the balance of reasons would remain in favor of abstention from alcohol. So the introduction of a rule excluding the reasons for drinking would not make a difference to the balance of reasons.

18. *Id.* at 47.

19. There is no corresponding ambiguity concerning 1 and 2. It is clear that in 1, Shapiro means the behavior the agent exhibits in the specific circumstances. Similarly, 2 refers to the belief of the agent that the rule applies in the specific circumstances.

20. Shapiro's argument would not hold if one assumed a specific type of fine-grained account of individuation of action according to which the motivation of an action is a defining property of it. For example, on this account, the action "behaviorally conforming to R (the self-imposed rule to abstain from alcohol) because the agent is motivated by R" is different from the action "behaviorally conforming to reason r (say, a desire to spend less than $2 on drinks) because the agent is motivated by r" (though in both cases the agent abstains from alcohol). On this

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

For example, on 3* my recovering alcoholic self must believe that in the specific circumstances I am facing (i.e., in that specific bar I am in, on a specific day and time, given my current psychological condition, i.e., my current beliefs, intentions, etc.) I *might* have ordered an alcoholic drink were it not for the rule. By contrast, on 3** it suffices that my recovering alcoholic self believes that there is at least one possible case in which absent the motivation of the rule I *might* have ordered an alcoholic drink (though she might also believe that in the specific circumstances I have a reason other than the rule that prevents me from ordering an alcoholic drink).

To put it more formally, the ambiguity affects the choice of the possible worlds that determine the truth value of the *might* counterfactual. On 3* the agent believes that the relevant closest possible world, P, *must* be one that differs from the actual world (condition C1) only in respect of one basic change in her psychological condition, that is, absence of motivation of R (as well as in respect of the minimal ramifications to C1 this absence entails). On 3* the *might* counterfactual is true if in P the agent fails both to be motivated by R and to conform to R. It is false if in P the agent both lacks the motivation of R and conforms to R. Condition 3** allows that the agent believes that the relevant closest possible world, P*, *might* be more distant from C1 than P and differ from C1 in more respects than simply the absence of R. On 3** the *might* counterfactual is true if in P* it is the case that both the agent is not motivated by R and does not conform to R. It is false if in P* the agent is not motivated by R and conforms to R.[21]

Condition 3 is what Raz calls a "detached" statement.[22] That is, it is a descriptive statement about the internal point of view of a belief-holder made by someone (say, a philosopher of action) who understands (though not necessarily accepts) this internal point of view. Were 3 to state only that absent the motivation of R the agent might not have conformed to R, it would be closer to the gist of Lewis's analysis of counterfactuals to take P to determine the truth value of the *might* counterfactual. In that case, 3 would have been a purely external statement that did not purport to describe the internal point of view of an agent. But 3 is a statement about the *belief* of an agent. The question is which of 3* and 3** best represents the content of the belief of that agent. For example, the question is how best to represent

fine-grained account, rules do not make a *practical* difference. They make a *logical* difference to the type of action one performs: since motivation partly defines action, it is logically trivial that absent the motivation of R, one *would not* conform to R. In general Shapiro seems to presuppose a coarse- rather than a fine-grained account of action individuation. For a coarse account, *see* Donald Davidson, *Action, Reasons and Causes, in* ESSAYS ON ACTION AND EVENTS 3 (1980); for a fine-grained account, *see* ALVIN GOLDMAN, A THEORY OF HUMAN ACTION (1970).

21. I follow (as Shapiro does) Lewis's account of counterfactuals; *see* D. LEWIS, COUNTERFACTUALS (1986). For the sake of simplicity, I presuppose the limit assumption and do not deal with questions concerning vacuously true counterfactuals. The meaning of 3* and 3** is determined by how the agent understands the motivation of and the conformity to R (i.e., the content of P or P*). As I later explain, the motivation of R should be understood in terms of folk psychology and the conformity to R in terms of folk ontology.

22. J. RAZ, THE CONCEPT OF A LEGAL SYSTEM, 236 (1980).

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

the judge's belief that absent the motivation of the rule of recognition she might not have conformed to it. Does the judge have a P-world or P*-world in mind? To this question Shapiro's account provides no answer. Thus 3 remains ambiguous between 3* and 3**.

The fact that 3 is a detached statement that purports to describe the internal point of view of the belief-holder has a further consequence. In order for 3 to represent faithfully the content of the relevant belief, it must respect the constraints of referential opacity. Given that the belief-holders are ordinary people, they cannot plausibly be assumed to employ anything other than the conceptual apparatus of folk psychology and ontology. But then 3 must provide a description of the content of the relevant belief which draws on nothing other than this conceptual apparatus. This means that 3 must describe the P-world or the P*-world (depending on how one disambiguates it) using nothing other than the conceptual resources of folk psychology and ontology.

II. THE PRACTICAL-DIFFERENCE THESIS AND THE DETERMINATION OF ACTION BY REASONS

Sometimes an agent has more than one reason inclining her toward a particular course of action but performs the action for only one of these reasons. I will illustrate this case with the following example. A godfather promises a gangster $50,000 if he kills a certain mafioso who happens to have murdered the gangster's brother. The gangster both needs the money and wants to take revenge for the death of his brother. He agrees to be the hit man and murders the mafioso. Later he says to a fellow gangster: "it was both 'business' and 'pleasure'; I would have done it even if it was only for the money but I really did it for revenge."

The gangster has offered a motivational explanation of his action. The gangster believes that he had two reasons for killing the mafioso: his desire for $50,000 and his desire for revenge. He further believes that absent the latter desire he would have still killed the mafioso. Finally, he believes that he killed him in order to take revenge and not in order to take the money.

The gangster's motivational explanation is of a special kind. It not only identifies the reasons inclining him toward the course of action he followed; it also singles out the reason why the action was performed. In an overall account of the psychological state the gangster was in while premeditating and executing the crime the desire for money has a place, and indeed an important one. But if one wants to find out what actually got the gangster to pull the trigger and shoot the mafioso, one needs to single out the desire for revenge. The latter desire is the reason why of the gangster's killing the mafioso. It is the reason that did not merely incline him toward but also

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

determined his action.[23] I will say that the gangster provided a motivational explanation in which he identified the *determinant* reason of his action.

An inclinatory reason is a reason that inclines one to act and could get one to act but in the specific circumstances does not get one to act. A determinant reason is a reason that in the specific circumstances gets one to act and can be cited as the reason why of one's action.[24] So we can define a determinant reason as the reason that not only inclines and could get the agent to act but also is the reason why of a performed action.[25]

The existence of determinant reasons allows us to speak of the phenomenon of the motivational *determination* of action by reasons. The determination of action may be understood either causally or teleologically, depending on one's preferred theory of motivation. So a determinant reason may be a reason that has in the appropriate way caused the action or else the end for which the action was done. It is irrelevant for the purposes of this paper whether the determination of action is understood in causal or teleological terms.[26]

Either of the two formulations of the practical-difference thesis fails to capture the phenomenon of action determination.[27] To show this, I will modify my initial example of the self-imposed rule to abstain from alcohol on medical grounds. Say that I have also adopted a second rule purely for financial reasons, namely, to avoid spending more than $4 on my evening outings. That is, the latter rule has independent grounding and has not been adopted as another precautionary measure against alcohol consumption. Assume further that I spend my evenings in a rather expensive bar where all alcoholic drinks cost more than $4. So before ordering a drink in that bar I know that I have two rules that prevent me from ordering an alcoholic

23. My account of the phenomenon that I call "determination of action by reasons" does not presuppose a *deterministic* theory of human action. It is compatible with my account that our free will selects the reasons for which we act. I use the notion of determination to denote the special relation between some motivating reasons and action. It does not imply anything about how the agent came to accept the corresponding reasons.

24. I owe the distinction between reasons that "incline us to act" and reasons that "get us to act" to JONATHAN DANCY, ETHICS WITHOUT PRINCIPLES 21 (2004). Note that I do not observe here the further distinctions Dancy makes between favorers, enablers, and intensifiers; *id.* at 38–52.

25. I am willing to allow that an agent who uses the conceptual apparatus of folk psychology may believe that a particular action of hers was overdetermined. If one operates within the conceptual framework of neurophysiology or physics, one may treat overdetermination as a rare phenomenon. But I have no theory about how rare a phenomenon folk psychology takes the overdetermination of action by reasons to be. However, even in cases of overdetermination of action, the distinction between determinant and inclinatory reasons may apply. Think, for example, of a case in which one has three reasons inclining oneself toward doing A, only two of which actually (over-)determine A. In this case the third reason would be the (merely) inclinatory reason.

26. For a recent defense of the causalist account of motivation, *see* MELE, *supra* note 2, at 38–60; for a teleological account, GEORGE WILSON, THE INTENTIONALITY OF HUMAN ACTION (1989).

27. For the sake of simplicity, in what follows, by "action determination" I mean "motivational determination of action."

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

beverage: my rule to abstain from alcohol and my rule to abstain from spending more than $4 in my evening outings.

Assume that I order a nonalcoholic beverage and that my determinant reason was the rule to abstain from alcohol rather than the financial rule. If we treat the practical-difference thesis as equivalent to condition 3*, I am not motivationally guided by that rule. For though I fulfill the first two conditions of the feasibility thesis (that is, I have conformed to the rule since I have ordered a nonalcoholic beverage and I believe that since the rule to abstain from alcohol applies in the circumstances, I do not have the option of ordering an alcoholic drink), I do not fulfill condition 3*. Since I know that I have also adopted another rule, namely, to abstain from spending more than $4 in my evening outings, which also applies in the circumstances and constrains me not to order an alcoholic drink, I cannot believe that absent the motivation of the rule to abstain from alcohol (on medical grounds), I might have ordered an alcoholic drink in the circumstances. So the rule to abstain from alcohol does not make the type of practical difference described in 3*. Since 3* is not fulfilled, I cannot be said, according to the feasibility thesis, to be motivationally guided by the rule to abstain from alcohol.

In fact 3* makes it in principle impossible for anyone who has adopted two different rules that in the circumstances incline one toward the same course of action to be motivationally guided by either of them. Say that X has adopted two rules, R_1 and R_2, which incline her toward doing A in circumstances C_1. According to 3*, in order for X to be motivationally guided by R_1, it is necessary that she believes that absent the motivation of R_1, she might not do A. But she cannot believe this since she knows that she has another motivation for doing A, namely, R_2. And she cannot be motivationally guided by R_2 either since she knows that absent the motivation of R_2, she would be motivated to do A by R_1.

On 3** I am motivationally guided by the rule to abstain from alcohol. For 3** requires only that the agent believe that the motivation of the rule makes a practical difference in at least one possible counterfactual scenario in which the rule applies, though not necessarily in the one the agent is currently facing. So, since there is a possible counterfactual case in which the rule to abstain from alcohol makes the required practical difference (say, one in which I have not adopted the financial rule), all three conditions of the feasibility thesis are satisfied. However, on 3** I can also be said to be motivationally guided by the financial rule, even though, as I have assumed, only the rule to abstain from alcohol has actually determined my ordering a nonalcoholic beverage. The rule to abstain from spending more than $2 in my evening outings requires that I order a nonalcoholic beverage at the pub. So the first condition of the feasibility thesis is fulfilled, since I conform to the rule. But so is the second condition. For since I have adopted the financial rule and recognize that the rule applies in the circumstances, I believe that nonconformity with the rule (i.e., ordering an alcoholic drink)

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

is infeasible. Finally, 3** is also fulfilled, since I believe that there is a possible counterfactual case in which the financial rule makes a practical difference (say, the one in which I have not adopted the rule to abstain from alcohol on medical grounds).

In general, if 3** holds, the agent may be said to be motivationally guided by almost every rule that applies in the circumstances and inclines the agent toward the relevant course of action she has actually performed. For if the agent's behavior in the circumstances is such that it conforms to more than one rule, and if the agent has adopted each of these rules that apply in the circumstances, then the first two conditions of the feasibility thesis hold for each of these rules. The third condition, 3**, is too weak to single out the rule that determined the action. Given that it is possible to imagine cases in which one might not perform the relevant action absent the motivation of almost any of these rules, almost every one of these rules will pass the test of 3**.

To sum up, either of the two readings of the practical-difference thesis fails to capture the phenomenon of action determination:

1) On 3* it is possible that a rule fails to satisfy the practical-difference thesis but determines action.
2) On 3** it is possible that a rule satisfies the practical-difference thesis but fails to determine action.

Shapiro's practical-difference thesis presupposes a counterfactual account of mental causation. The counterexample I have produced against 3* resembles cases of trumping preemption; it is in fact a case of mental trumping preemption.[28] Can Shapiro's practical-difference thesis understood as 3* be rescued if we adopt any of the strategies proposed by supporters of the counterfactual account of causation to deal with trumping preemption?[29]

The first strategy is to deny that mental trumping preemption exists at all. It could be argued that there is a relevant "cutting" of the causal chain of R2 (the financial rule) to action A (ordering a nonalcoholic beverage) at the neurophysiological level. For example, a signal, S1, coming from a neuron, N1, activated by R1 (the rule to abstain from alcohol on medical grounds) may both activate a neuron, N2, responsible for the initiation of A and block off a signal, S2, to N4, to similarly initiate A coming from the neuron, N3, which is activated by R2. In this case the counterfactual account could still identify R1 as the determinant reason for the action. The counterfactual scenario "absent the motivation of R1" would be a scenario in which N2 is

28. For an account of (nonmental) trumping preemption, *see* J. Shaffer, *Trumping Preemption, in* CAUSATION AND COUNTERFACTUALS 59 (J. Collins, N. Hall & L. A. Paul eds., 2004). I discuss the relation between mental and nonmental trumping preemption in my *Mental Trumping Preemption* (unpublished manuscript).

29. For the description of the three strategies I draw freely from D. Lewis, *Causation as Influence, in* CAUSATION AND COUNTERFACTUALS 75 (J. Collins, N. Hall & L. A. Paul eds., 2004). Lewis mentions all three strategies but explicitly endorses only the third.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

not activated. But then A might not have occurred because on this scenario N4 also remains inactive (due to its being blocked by N1).

The second strategy is to treat action A as a "fragile" event assuming a fine-grained account of individuation of action. On this scenario the action, A*, caused by R2 would differ in some respects from A caused by R1. Think, for example, that there are certain regularities governing how motivation is expressed in behavior. Assume that when I am motivated by the financial rule, the tone of my voice when asking for a nonalcoholic beverage or my facial expressions are slightly different from when I am motivated by the medical rule. If this is so, then absent the motivation of R1 I might not do A but rather the slightly different action A*.

The third strategy is to adopt Lewis's account of causal influence. On this account not only the events caused but also their causes are "fragile." Lewis invites us to think of a range of alterations (C_1, C_2, $C_3 \ldots C_n$) of an actual cause C and map them onto a range of alterations (E_1, E_2, $E_3 \ldots E_n$) of the caused event E so as to create a pattern of counterfactual dependence (if C_1 had occurred E_1 would have occurred, if C_2 had occurred E_2 would have occurred and so on). C causally influences E if and only if there is such a pattern of counterfactual dependence. In our example, we are invited to think of a range of similar alterations of R1 (say, $R1_1$: "abstain from alcohol and urge the barman to stop serving alcohol") and map them onto similar alterations of A (say A_1: "order a nonalcoholic beverage and urge the barman to stop serving alcohol"), creating a pattern of counterfactual dependence while holding R2 fixed. In this case if an alteration of R1, $R1_1$, had occurred, then an alteration of A, A_1, would have occurred. By contrast, if we keep R1 fixed and make alterations to R2, then there is no difference to action A. In this case the *might* counterfactual "absent the motivation of R1, the agent might not conform to R1" should be considered equivalent to "absent the motivation of R1 understood as a fragile cause, the agent might not conform to R1 equally understood as a fragile event."

Even if we assume that any of these strategies might help the counterfactual account of causation tackle cases of trumping preemption, none of them is of avail to the practical-difference thesis. For, as I have already indicated, the practical-difference thesis is a detached statement about a) the beliefs of the agent from b) her internal point of view c) which draws exclusively on the conceptual resources of folk psychology and ontology. But each of these strategies requires the agent to have beliefs drawing from resources of more refined conceptual apparatuses.

The first strategy requires that the agent possesses a neurophysiological account of mental causation. The second strategy requires that the agent espouse a specific fine-grained account of individuation of actions. The third strategy requires that the agent accept in addition an account of the fragility of causes. These metaphysical positions may be invoked in philosophical discussions of causation but should not be invoked in an analysis of the content of a belief of an ordinary agent (i.e., a judge or a citizen) about

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

the practical difference the rules make that aims to represent her internal point of view. An analysis of the content of such a belief of an ordinary agent should use the conceptual apparatus of folk psychology and ontology and not of neurophysiology and analytic metaphysics if it is to be faithful to the internal perspective of the belief-holder.

So if the detached statement of the practical-difference thesis is to represent faithfully the content of the rule-follower's belief, it cannot equate the motivation of the agent with a neurophysiological condition. Rather the belief that the rule applies and the motivation to conform to the rule should be understood as folk-psychological conditions. Neither should conformity with the rule be understood from the internal point of view as a fine-grained or fragile action, as the second and the third strategy require. To treat something as a fine-grained or fragile action is similar to treating it as a finely individuated token of a certain action-type. On such an understanding, for example, a1, ordering a nonalcoholic beverage at the bar at approximately 9 PM on February 1, 2006, is a different action from a2, ordering a nonalcoholic beverage at the bar a second later, though both a1 and a2 may be tokens of the same action-type, A, that is, ordering a nonalcoholic beverage at the bar. Now, it is implausible to assume that someone who adopts the rule to abstain from alcohol on medical grounds wants to make certain that she will do a1 rather than a2. Instead of being concerned with a specific finely individuated token of the action-type A, she wants to make certain that she does any relevant token of A as opposed to any token of the action-type B, that is, order an alcoholic drink at the bar. So, from the internal point of view of the rule-follower (which the detached practical-difference thesis purports to capture), "conformity to the rule" must be understood to refer to types of actions as opposed to finely individuated tokens (a justification of the individuation of which may require recourse to analytic metaphysics). For example, what the agent believes when she believes that absent the motivation of the rule to abstain from alcohol she might not conform to the rule is that she might not do the action-type of ordering a nonalcoholic beverage at the bar (or else *any* relevant token of this action-type) and not that she might not do the specific finely individuated token of ordering a nonalcoholic beverage at exactly 9 PM on February 1, 2006.[30]

30. The following clarifications are needed: 1) I do not claim that the ordinary rule-follower is unable to differentiate finely individuated tokens of the same action-type. My point is rather that she thinks of conformity with the rule in terms of action-types rather than in terms of finely individuated action-tokens. 2) Different action-types may count as conformity to the same rule R (say, the rule to abstain from alcohol), for example, "order a nonalcoholic drink at a bar," or "refuse to drink during a meal." For the sake of simplicity, I associate in the main text conformity to a rule with only one action-type. 3) There is a further question of how an ordinary rule-follower individuates action-types that count as conformity to a rule. I cannot deal in this paper with action-individuation in folk ontology. It suffices for my purposes to show that the ordinary rule-follower does not understand conformity with the rule as a fine-grained action in the way required by the analytic metaphysics employed by those who defend counterfactual accounts of causation.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

Thus the proper analysis of the practical-difference thesis as a detached statement is the following: the rule-follower believes that absent the motivation of the rule understood as a folk-psychological condition, she might not do (any relevant, no matter how finely individuated, tokens of) a certain action-type A that counts as conformity to the rule. This analysis shows that any of the strategies available to a defender of counterfactual accounts of causation against the challenge of trumping preemption are not available to the defender of the practical-difference thesis. For it is compatible with the practical-difference thesis thus captured in terms of folk psychology and ontology that the agent might also believe that she has another motivation from a different rule (similarly understood as a folk-psychological condition) that might get her to do (any, no matter how finely individuated, tokens of) the same action-type A as the original rule.[31] In this case (any finely individuated tokens of) the same action-type A would count as conformity to both rules. The challenge remains that if restricted to the conceptual apparatus of folk psychology and ontology (as one should be, given that the practical-difference thesis purports to describe the internal point of view of the rule-follower) one cannot square the practical-difference thesis with the determination of action by reasons.

The fact that the practical-difference thesis is a detached statement purporting to describe the internal point of view of the rule-follower helps us to counter the following objection.[32] I have claimed that in the case in which the practical-difference thesis is read as 3*, the rule-follower who has adopted both R1 and R2 cannot believe, B1, that absent the motivation of R1 she might not do A because she also believes, B2, that either one of R1 or R2 on its own suffices to get her to do A. However, goes the objection, if any of the strategies defending a counterfactual account of causation against the challenge of trumping preemption works, B1 and B2 are consistent, and the rule-follower can hold them both without any loss of rationality. True, the agent might not be able to explain their consistency, but the fact that they are consistent suffices to exonerate her from any charge of irrationality.

The objection requires that B1 and B2 are understood as follows:

B1*: Absent the motivation of R1, I might not have done the (finely individuated) token a1 of the type-action A.

B2*: Either one of R1 or R2 on its own suffices to get me to do a certain (finely individuated) token of the same type-action A, though not the same token (say, R1 would get me to do a1 and R2 would get me to do a2).

31. More precisely, I have in mind someone who knows that a) R1 (the rule to abstain from alcohol) suffices to get her to do (action-type) A (order a nonalcoholic beverage at the bar); and b) R2 (the rule to avoid spending more than $4 in the evening outings) suffices to get her to do (action-type) B (order a drink that costs less than $4) and infers that in the specific circumstances, to do B is to do A (any token of A is also a token of B).

32. I am grateful to an anonymous referee for bringing this objection to my attention and forcing me to clarify the nature of the practical-difference thesis as a detached statement.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

B1* and B2* are consistent, and provided that counterfactual accounts of causation can handle mental trumping preemption, it would be perfectly rational for an agent to hold them both. However, the person who holds B1* and B2* seems to adopt a rule in order to secure that she does a specific (finely individuated) token of a certain action-type as opposed to doing (any relevant finely individuated tokens of) a certain action-type. But as I argue above, it is implausible to assume that this is the end of an ordinary rule-follower. I suggest that if we are best to capture the internal perspective of the ordinary rule-follower, the content of B1 and B2 should be analyzed as follows:

B1** Absent the motivation of R1, I might not have done (any of the finely individuated tokens of) the action-type A.

B2** Either one of R1 or R2 on its own suffices to get me to do (a certain finely individuated token of) the same action-type A.

B1** and B2** are inconsistent. The agent cannot hold them both without loss of rationality. And if I am right, B1** and B2** best represent the internal point of view of the rule-follower.

To put it differently, the objection is based on a transparent reading of the beliefs of the rule-follower and presupposes a fine-grained account of action-individuation. The practical-difference thesis, however, is a detached statement purporting to describe the internal point of view of the rule-follower, who cannot be assumed to employ anything other than the conceptual apparatus of folk psychology and ontology. So the description it provides must get as close as possible to meeting the constraints of referential opacity.

To recapitulate, my argument against the practical-difference thesis is not that it presupposes a counterfactual account of causation and that a counterfactual account cannot handle mental trumping preemption. My argument is rather that since the practical-difference thesis is a detached thesis about the beliefs of agents who employ the conceptual framework of folk psychology and ontology, the defender of the practical-difference thesis has to tackle cases of mental trumping preemption within the conceptual framework of folk psychology and ontology. But the proposed modifications to the counterfactual account of causation so that it handles trumping preemption require one to step out of this framework.

To do justice to action determination, a possible principle of motivational guidance of rules should include the following condition:

> Rule R prescribing that X does A motivationally guides X in doing A only if X does A and R determines X's doing A.[33]

33. R's determination may be understood in either causal (as, for example, in the context of a counterfactual theory of causation) or teleological terms.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

But can rules determine actions? So far I have been assuming that rules provide motivational guidance and I have been trying to assess whether the practical-difference thesis provides a necessary condition for the motivational guidance of rules. In the remainder of the paper, I examine whether rules *can* motivationally guide. The revised principle of motivational guidance suggests that this question is equivalent to the question of whether rules can determine action. I suggest that rules qua rules cannot determine action. In order to substantiate this conclusion, I focus on Shapiro's interpretation of Hart's account of rules as content-independent and peremptory reasons for action.

III. RULES AS CONTENT-INDEPENDENT AND PEREMPTORY REASONS FOR ACTION

On Shapiro's interpretation, for Hart "a reason is content-independent if the validity of the reason is independent of the validity of its content. Rules are content-independent reasons in that agents who are committed to the rule try to conform simply because the rule regulates the action in question."[34] For example, the fact that I have made a promise to do A is a reason for doing A independently of the merits of A itself. Rules are content-independent reasons for action because the fact that a rule prescribes a certain action is a reason for doing it independently of the merits of the action. Shapiro translates the normative character of content-independent reasons into a kind of motivational power. To treat a rule as a content-independent reason for action is to conform to it simply because it is a rule.

A peremptory reason is not a reason for doing a particular action. It is rather a reason not to deliberate on the merits of the action. According to Shapiro, for Hart:

> rules are "peremptory" reasons for action, i.e. reasons not to deliberate on the merits of the case at hand. When someone who is guided by a legal rule is confronted with a situation where the rule applies, that person does not deliberate about whether to follow the rule—he or she simply follows the rule.[35]

This account of the peremptoriness of rules accords with the second condition of Shapiro's feasibility thesis. Rules render choices infeasible by being peremptory reasons for action. And it provides some support for the problematic, as I have argued, practical-difference thesis:

34. Shapiro, *On Hart's Way Out*, *supra* note 1, at 492–493.
35. *Id.* at 493.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

> When rules are seen as peremptory reasons, it follows that people can take them in these ways only when they believe the rules are capable of making practical differences. For if rules gave the same answers as deliberation on the merits, agents would never believe themselves to have reasons not to deliberate about a rule's recommendations.[36]

So rules motivationally guide action by being content-independent and peremptory reasons for action. The agent is constrained to conform behaviorally to the rule because she treats it both as a content-independent reason (she conforms "simply because" she thinks that the rule applies) and as a peremptory reason (she does not believe she has a choice). Further, since she treats it as a peremptory reason, the rule makes also a practical difference.

As I argue in the previous section, the practical-difference thesis is problematic, and a proper analysis of motivational guidance must capture the phenomenon of action determination. So if, as Shapiro claims, rules qua rules motivationally guide action, peremptory and content-independent reasons must be determinant reasons. Are they?

I begin by elaborating on my previous distinction between inclinatory and determinant reasons. First, there is a distinction between considerations that incline us to perform a certain action, A, considerations that incline us against doing A, and considerations that do not incline us either toward or against doing A.[37] A consideration inclines one to do A only if one takes it to be a reason for one's preferring to do A. For example, the gangster's desire for revenge inclines him toward killing the mafioso. His possible belief that murder is morally bad inclines him toward not killing the mafioso. His belief that he shoots best if he has a good night's sleep the day before the killing inclines him neither toward nor against killing the mafioso. It may incline him, for example, toward not spending the night at a nightclub. Second, some of the reasons that incline one to act would suffice to get one to act but do not in the specific circumstances get one to act. These are the *merely* inclinatory reasons in the circumstances. By contrast, the reason that actually gets one to act in the circumstances is the determinant reason or the reason why of one's action. In our example, the gangster's desire for revenge is the reason why he killed the mafioso, while his desire for money is a merely inclinatory reason. On the basis of these distinctions it follows that the determinant reasons of doing A are a subclass of the inclinatory reasons to do A.[38]

36. *Id.* at 494.

37. I refer here to action-types.

38. This is by no means a comprehensive list of reasons for A. For example, I have not discussed whether there are reasons that incline us to act but can never get us to act (i.e., are necessarily nondeterminant). This discussion is far beyond the scope of this paper. I focus only on distinctions that are immediately relevant to my argument.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

It is clear that rules cannot determine action in virtue of the fact that they are peremptory reasons for action. For a peremptory reason for A does not even incline the agent toward doing A. It inclines the agent toward *nondeliberation* on the merits of A.[39] If it does not incline one to do A, it cannot be a determinant reason for A. So rules cannot motivationally guide the agent in virtue of their peremptoriness.

Shapiro puts a lot of weight on the psychological effects of the peremptory character of rules. When probing a speculative account of the psychological roots of the constraint model, he singles out the psychological power of repression. We can say that repression is for Shapiro the psychological counterpart of peremptoriness. This is how he takes rules to repress the will:

> First, the rule might prevent the agent from considering the reasons for disobeying the rule. The reasons for breaking the rule would, in other words, be "repressed." . . . Second, the rule might disable the person's normal psychological inhibitions. An agent in the grip of a rule might no longer be able to withstand certain emotional pressures, such as guilt and shame, even though the agent would be aware of the important reasons for breaking the rule.[40]

Even if we accept this rather extreme thesis about the repressive powers of rules, it still does not follow that rules qua peremptory reasons incline one toward conforming to the rule. At most they can be said to incline one toward blocking off thoughts about the possibility of nonconformity, *were one for some other reasons inclined to conform to the rule.* Some of the latter (presumably content-independent) reasons, and not the peremptory reason, could be the determinant reasons.[41]

The remaining possibility is that rules motivationally determine the actions of the agent in virtue of their being content-independent reasons for action. Unfortunately, Shapiro in his articles focuses almost exclusively on the peremptory character of rules and does not elaborate on how rules qua content-independent reasons may be thought to affect the motivation of the rule-follower. For this reason I resort to a rather speculative reconstruction of Shapiro's argument. I start with some preliminary distinctions. First, I distinguish rule-following activity from mere habitual behavior. The

39. Of course, one *consequence* of one's treating something as a peremptory reason might be that one will ultimately get to act in the required way. But this will not be the *result* of one's treating something as a peremptory reason. For a relevant distinction between results and consequences, *see* G. H. VON WRIGHT, NORM AND ACTION 39–41 (1963).

40. Shapiro, *The Difference That Rules Make*, *supra* note 1, at 53.

41. On Shapiro's interpretation, Hart treats peremptory reasons as blocking deliberation altogether. As I have already stated, I do not examine the exegetical value of Shapiro's interpretation in this paper. Further, though Hart's concept of peremptory reasons was heavily influenced by Raz's account of exclusionary reasons, in this paper I do not intend to address the issue of whether exclusionary reasons as Raz understands them are motivationally inert. Neither am I concerned with the issue of whether, on Raz's account, rules as protected or preemptive reasons for action rules are motivationally inert.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

minimal requirement the agent must satisfy to engage in a rule-following activity is to be able to identify something as the rule she is following *upon reflection* on her practice. So, for example, if I follow the self-imposed rule of abstaining from alcohol, I should be able to identify it as the rule I am following upon reflecting on my behavior. Mere habitual behavior need not satisfy such a condition. For example, if I am in the habit of listening to the radio broadcast of football matches every Sunday, I may not identify this behavior as a rule I follow. Second, I will distinguish between a) the phase of the acceptance of the rule, and b) the phase of the application of the rule.[42] Third, either one of a or b may be deliberative or nondeliberative. For example, I might have adopted a rule to abstain from alcohol after deliberating about the effects of alcohol consumption on my health. Or I might have adopted the same rule simply on the advice of my doctor.

Further, the application of the rule might be itself the result of a further deliberation. For example, while in a bar with a friend I might deliberate as follows: "I have adopted a rule to abstain from alcohol. So I should order lemonade." Alternatively, the agent might have internalized the rule and respond "automatically" to possible cases of application of the rule. A driver who has "internalized" the rule to buckle his seat belt might do so upon entering his car without going through any deliberative steps concerning the application of the rule.[43]

Finally, a statement of the type "X follows an instrumental rule R" might be read in two ways. It might be a statement made from an external perspective. In this case, it asserts that X follows R and R is an instrumental rule. Alternatively, it might be a (detached) statement about the internal perspective of the agent. In the latter case, it asserts that X follows R and believes that R is an instrumental rule. On the external perspective, it does not matter whether X believes that R is an instrumental rule. On the internal perspective, it does not matter whether R is actually an instrumental rule. It is necessary that X believes so. But as I explain above, it is not necessary that X is aware of this belief at the time of conforming to R. It suffices that X may become aware of her belief upon reflection.

Since I am concerned in this paper with the motivation of the rule-follower, I focus on the internal perspective. The task is to explain the motivation of the agent who (at least upon reflection) is able to identify

42. I believe it is important to keep separate the considerations that incline us or get us to accept a rule from those that incline us or get us to conform to a rule. For example, on this distinction, considerations concerning the legitimacy of the rule belong to the former considerations. This is why I take the belief that a certain rule is a legitimate standard of conduct that Shapiro claims the rule-follower acts upon to determine the acceptance of rather than the conformity with the rule. It affects the process of conforming to the rule indirectly by being the grounds of an intention or a desire to conform to the rule (*see* my subsequent account of the motivational explanation of deliberative conformity to an instrumental rule).

43. Note that the difference of the rule-follower agent who buckles her seat belt without deliberating about it and someone who buckles her seat belt out of mere habit is that only the former can identify, upon reflection, a relevant rule that governs her action.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

something as a rule she is following and believes that it is an instrumental rule. I start by examining the case of deliberative application of an instrumental rule that is based on deliberative acceptance of that rule.

After deliberation, I adopt the rule to abstain from alcohol for the rest of my life. The main reason that led me to this decision is my concern for my health. I conceive this rule as instrumental to a healthy life. Further, let us assume that I consider that there might be certain exceptional situations in which I might decide not to apply my self-adopted rule. So I have adopted a rule to abstain from alcohol (rule R) in nonexceptional circumstances C.

Now, I enter into a bar and my rule comes to my mind. I form the intention to order a nonalcoholic beverage. How is the formation of this intention to be explained? Think of the following motivational explanation:

(ME1) (I1) I intend to conform to R (abstain from alcohol) in C.
(B1) I believe that I am in C.
(B2) I believe that R requires that I do A (order a nonalcoholic beverage) in C.
So
(I2) I form the intention to do A.[44]

In ME1, R qua content-independent reason for action has no motivational role. Let us couch ME1 in causal terms. Though R is the content of I1, it is my intending in I1 in conjunction with my believing in B1 and B2 which moves me to I2. R has only an epistemic function. I learn from R that I should do A (B2). This seems to be the only thing R does as a content-independent reason for action. I do not have to deliberate about the pros and cons of my doing A. I can rely on the authority R has qua rule. It informs me that there are sufficient and overriding reasons supporting my doing A. But I do not form the intention to do A because R has any motivational power qua content-independent reason, as Shapiro's account requires.

To put it differently while reading ME1 with my teleological glasses on, the end of my intention I2 is to follow the content of my prior intention I1, which is my conformity with R. It is not to follow the rule qua rule, that is, qua content-independent reason for action.

It could be objected that for R to motivate qua content-independent reason is for I1 (the intention to conform to R) to get one to perform a specific action that counts as conformity to R. But this cannot capture the notion of motivational guidance by a rule Shapiro has in mind. For when I order a nonalcoholic beverage and thus conform to R in order to please my girlfriend, who wants me to adhere to R, I may still be motivated by I1. In this case, my desire to please my girlfriend gets me to form I1, and I1

44. In, Bn, and Dn (later in the text) stand for the states of intending, believing, and desiring and not for the content of the relevant intentions, beliefs, and desires.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

gets me to form I2.[45] But clearly this case is not one in which R gets me to conform to it in Shapiro's sense. In this case I conform to R because I want to please my girlfriend and not "because of the rule." R again seems to provide nothing other than (mere) epistemic guidance. So it is a mistake to equate the motivational clout of R qua content-independent reason with the motivational clout of I1.

What would it be to be motivated by the rule qua content-independent reason for action and not qua the content of a former intention? Shapiro claims that when an agent treats a rule as a content-independent reason, she is motivated to conform with it because it is a rule. So we can think of the rule-follower as having an additional desire, D1, to conform to *a* rule and introduce D1 in our motivational explanation. On this new motivational explanation, ME2, what determines the formation of my intention to do A is not (or, at least, is not simply) my intending in I1; it is my desiring in D1. Alternatively, the end of my I2 is not to follow the content of my intention I1 but to follow the content of my desire D1. I do A in order to conform to *a* rule.[46]

I believe that the introduction of D1 provides an alternative way of understanding Shapiro's claim that on the constraint model the rule-follower believes that nonconformity with the rule is infeasible. It is not B2 that makes the agent believe that she does not have the option of nonconformity as Shapiro's formulation of the feasibility thesis seems to suggest (indeed, how could a belief like B2 have this effect?);[47] it is rather the strength of D1 (in conjunction with B2). D1 must be assumed to be the strongest desire the agent has for conforming with the rule once she adopts the rule (that is, stronger than all other desires that incline her toward nonconformity). Since the agent holds B2, and D1 is her strongest relevant desire, she comes to believe that it is psychologically impossible for her not to conform.

I1 may be considered to activate D1. On this speculative reconstruction of Shapiro's account of the psychology of the rule-follower, once the agent

45. A similar case is that of a judge who applies the rule of recognition in court in order to keep her salary. In such a case, a judge may act out of an intention to conform to the rule of recognition but is not motivationally guided by it in Shapiro's sense.

46. In ME2, I1 can be understood as a background reason that inclines me toward A (and that, as I explain in the text, can activate D1). It informs me that I have strong reasons to prefer A and, assuming that every intention includes a desire, it makes me desire to conform to R in C. Perhaps in the absence of D1, I1 (in conjunction with B2 and B3) would get me to do A (as in ME1). But as it is, D1 and not I1 gets me to do A.

47. Shapiro's formulation of the second tenet of the feasibility thesis [this paper: p. 321] is obscure. It could be taken to mean that the belief that R applies makes it impossible that the agent does not conform to R. This would be an absurd claim as it would preclude that the agent might be forced by someone else not to comply with R. Shapiro obviously means that the belief creates a psychological constraint and the agent cannot "intentionally" fail to comply (*see* his later remarks in Shapiro, *Hart's Way Out, supra* note 1, at 51–54). In the text I try to provide a possible account of the workings of such psychological constraint. Because of the strength of D1 and of her belief that R applies, the agent comes to believe that she cannot intentionally fail to comply. This is an alternative account to Shapiro's own explanation in terms of repression of reasons I have already mentioned.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

adopts a certain rule, D1 becomes activated. The activation of D1 is at the heart of the psychological mechanism that makes the agent believe that nonconformity is an option for her and constrains her behavior. Statements of the type "X did A because X treated R as a content-independent reason for doing A" are meant to reflect this psychological component. On ME2, to treat R as a content-independent reason for action is to desire to apply *a* rule.

ME2 is problematic. Instrumental rules are hypothetical imperatives (e.g., "if you want to remain healthy, you should abstain from alcohol"). They are contrasted with rules that are categorical imperatives. The paradigmatic cases of categorical imperatives are moral rules (e.g., "you ought not to kill"). Rules are categorical because they have intrinsic value, that is, they are valuable for their own sake. So conformity with a categorical rule (e.g., not killing) is valuable for its own sake. By contrast, instrumental rules have only instrumental value. Conformity with an instrumental rule is valuable only as a means to a further end. This difference in value explains why a rule cannot be both instrumental and categorical. Further, if one desires to conform to a rule that one considers categorical, one should desire to conform to the rule for the sake of conforming to the rule and not for any other reason (provided that one is rational). In a similar vein, if one desires to conform to a rule that one considers instrumental, one should desire to conform for the sake of a further end and not for the sake of conforming to the rule.

In my example, I form I1 after deliberation about the instrumental value of R. And if desires are at least parts of intentions, then in having I1 I have an instrumental or extrinsic desire. I desire to conform to R as a means to health and not for its own sake. By contrast, D1 is not an instrumental desire. I do not desire to conform to a rule as a means to something else but rather for its own sake. So D1 is an intrinsic desire.

So if I have D1, then I must be committed to believing (at least implicitly) that the essential properties of a rule are valuable for their own sake. And since R is a rule, then I must believe that R must be valuable for its own sake. But if I believe that R is valuable for its own sake, then I am committed to believing that R is categorical. So on ME2, I have contradictory beliefs concerning the status of R. I believe both that R is instrumental (on the basis of I1) and that it is categorical (on the basis of D1).

One could protest that it does not follow that if rules are valuable for their own sake and R is a rule, R is a categorical rule. Rather it follows that it is categorical that I treat the rule as a normative requirement.[48] For example, in the case of the instrumental rule to abstain from alcohol, "it is categorical (*if I want to remain healthy, I abstain from alcohol*)." "Categorical" has a wide scope: it governs the whole conditional, not just the consequent

48. For an account of normative requirements, *see* John Broome, *Normative Requirements*, 12 Ratio 398 (1999).

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

of the instrumental rule. By contrast, in a categorical rule the modality has a narrow scope: it is categorical that one not kill.

There are two possible readings of D1. On the first, we can take the desire to conform with *a* rule to be equivalent to a desire to treat instrumental rules as normative requirements. This is precisely the way the objection suggests D1 to be understood. However, if this is how D1 is understood, then it will fail to have any motivational impact on the formation of I2. The motivational explanation provided would be a variant of ME1. The desire to follow the relevant normative requirement would not be the reason why of the formation of the intention to do A. It would simply underline the epistemic function of the rule.

On the second reading, D1 is a desire to conform to what a rule prescribes and not a desire to treat a rule as a normative requirement. As I explain above, this allows D1 to have a motivational impact on I2. D1 may figure as the reason why I form the intention to do A. But it has the drawback of requiring the agent to treat R as both instrumental and categorical.

One might suggest that the problem of crediting the agent with contradictory beliefs is eased by assuming that Shapiro actually favors what can be called a "modular" account. Once one forms I1 (and thus moves from the process of accepting the rule to the process of applying the rule), one shifts to a specific module in which the rule is conceived as categorical rather than instrumental. Previous considerations concerning the instrumental value of the rule are simply left out of the module.[49]

However, this modular account is problematic. First, it is reasonable to assume that considerations that would incline one against conformity with the rule are excluded. But why should a belief concerning the instrumental value of the rule (or, as I explain below, a desire for the relevant ends, conformity with which the rule promotes) be left out? This belief favors conformity rather than being a threat to it.

There are, of course, cases in which becoming oblivious to reasons for adopting an instrumental rule enhances the possibilities of achieving the desired end.[50] For example, perhaps the best way of achieving maximal pleasure in one's life as a whole is for one to accept certain rules that one treats as categorical and to cease considering that one's final end is pleasure. In such cases one cannot be said to be motivated by a desire for maximal pleasure in conforming to the rule, since, as per assumption, part of what one commits oneself to in accepting the rule is not to be motivated by hedonistic considerations. Though this is a plausible scenario, it cannot be said that all cases of instrumental rule-following are like that. In many cases

49. This modular account explores another possible understanding of the psychological function of the peremptoriness of rules. The rules are peremptory because once you adopt them, you start treating them as categorical. Considerations about their instrumental value are ignored.

50. I am grateful to Kent Hurtig for bringing the possibility of such cases to my attention.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

the agent remains aware of the instrumental value of the rule she follows. So the modular account does not cover all cases of instrumental rule-following.

Furthermore, the modular account covers "pathological" and not "paradigmatic" cases of instrumental rule-following since it is based on self-deception. The agent decides to endorse a policy of self-deception because apparently she is psychologically incapable of achieving her preferred ends if she is aware of these ends while trying to achieve them. Thus these cases belong to the "pathology" of instrumental rule-following. For this reason they may be better considered peripheral and not central cases of instrumental rule-following.[51] So the modular model cannot be the "standard" model of instrumental rule-following.

Second, in many cases that we have every reason to describe as genuine cases of rule-following, the agent might be motivated by a desire to achieve the ends the instrumental rule is supposed to serve. For example, when asked why I buckle my seat belt when getting into the car, I might respond by saying that conforming to this rule makes me feel safe. In this case of deliberative conformity with an instrumental rule, the agent appears to be fully aware of its instrumental value and in fact is motivated by considerations concerning the ends for which the rule was accepted in the first place. In contrast to what the modular account suggests, these considerations seem to determine the agent's conformity with the relevant rule.

This last example illustrates how one can avoid both ascribing an obvious contradiction in the beliefs of the rule-follower and endorsing the modular account as the "standard" model of instrumental rule-following. In the case of deliberative conformity with an instrumental rule, the agent's conformity may be motivated by considerations of the ends for which the rule was endorsed at the phase of accepting the rule. So we can come up with a new motivational explanation, ME3, by adding to ME1 instead of D1: a) a belief, B3, that conformity to the rule promotes the ends for which the rule was accepted; and b) a desire, D2, that these ends be achieved. In the case of my self-adopted rule to abstain from alcohol, B3 would be the belief that the rule of abstaining from alcohol is instrumental to my remaining healthy, and D2 the desire to remain healthy. I would suggest that it is desire for this end in conjunction with B1, B2, and B3 that determines my conformity to the rule (my intention to order a nonalcoholic beverage). On ME3, the considerations about the value of an instrumental rule that determine the agent's acceptance of the rule also determine the agent's conformity to the rule.[52]

51. I rely here on the distinction between "central" and "peripheral" cases supported by JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 9–11 (1980). Given that there are cases in which the agent can follow the rule without relying on self-deception, and given that self-deception is a device for overcoming a psychological defection, I think I am justified in treating cases of rule-following that rely on self-deception as peripheral. In any case, I think the burden of proof is on those who would like to treat such cases as paradigmatic.

52. In ME3, I1 functions as a kind of background reason having a peremptory function: it blocks off irrelevant considerations. The choice between ME1 and ME3 depends partly on

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

On ME3 (and on ME1 alike) the rule provides only epistemic guidance. It informs the agent about both the general policy she should follow and the specific action she should choose in order to achieve her desired ends. But it does not provide motivational guidance in the sense of motivationally determining her action *qua* content-independent reason for action.

ME3 is consistent with the peremptory function of rules. It does not require that the agent has to deliberate afresh about the desirability of the ends that conformity with the rule serves or about whether the rule is a satisfying means to these ends. It requires only that the agent is aware of the ends and that she desires that they be achieved.[53]

A further question is what happens in cases of nondeliberative conformity with an instrumental rule. I think we should distinguish at least two cases. In the first, the agent has an intention to conform to a rule that she is able upon reflection to identify as an instrumental rule but she is at the process of conforming to it unaware of the fact that she accepted it as an instrumental rule at a previous time. Think, for example, of the case in which before starting my car, I run through in my head a kind of checklist of several things I should do: "buckle my seatbelt," "adjust the side mirrors," "check the gas indicator," and so on. At the time I am not aware that these rules are conducive to safe driving but on reflection I can identify them as instrumental rules that I accepted either after independent deliberation or because someone else told me so, and so on. Am I in this case conforming to the rule because it is a content-independent reason for action?

I think not, for reasons similar to the ones that explain why in ME1 R is not a determinant reason. In ME1 the formation of my intention I2 was explained exclusively in terms of the motivational import of states of believing (B1 and B2) and intending (I1). Alternatively, the end of I2 is to fulfill a prior intention, I1, which happens to have R as its content. In a similar manner, I intend to conform to the rule "buckle my seat belt when entering the car" and this leads me to form the intention to buckle my seat belt. The difference is that states of believing similar to B1 and B2 do not intervene, since it is assumed that I form the intention to buckle my seat belt without going through any steps of deliberation. But again I find no

one's theory about the role of intentions in deliberation and about whether they constitute reasons for action. I cannot deal with these issues in this paper.

53. I have argued that on ME1 and ME3 instrumental rules qua content-independent reasons have an epistemic function but fail to be determinant reasons. They help the agent identify the course of action she has to perform if she is to conform to the rule but do not get her to act. A further question may be raised. Do rules qua content-independent reasons fail to be determinant reasons because they fail to be inclinatory reasons? Or is there a possible (rather loose) understanding of inclinatory reasons on which rules qua content-independent reasons incline the agent toward the performance of the required course of action but fall short of getting her to act? On this account, rules may be understood as *necessarily* merely inclinatory reasons. A full answer to this question would require an extensive analysis of the nature of inclinatory reasons, which is beyond the scope of this paper. So I prefer to leave this issue open, especially since nothing important hinges on it insofar as my main argument is concerned.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

room in this example for rules motivationally determining conformity qua content-independent reasons for action. They have only an epistemic role: through extensive repetition the agent has associated firmly in her mind the rule of buckling her seat belt with the specific action of buckling her seat belt so that there is no need to draw in her mind the relevant inference explicitly.

In the second case, the agent conforms to a rule that she is unable to identify as instrumental even upon reflection. She has internalized the rule but she does not remember any more why she has done so. But then she does not fulfill the minimal condition for being considered a follower of an instrumental rule, since from her internal perspective the rule is not instrumental. So her case is not a case of instrumental rule-following.

So far, I have identified two plausible motivational explanations of instrumental rule-following activity, ME1 and ME3, in both of which rules qua content-independent reasons have an exclusively epistemic function. It is now time to compare ME1 and ME3. D2, which figures explicitly only in ME3, may be thought to be implicit in ME1 as well. D2 was part of what motivated me to form I1.[54] So in a more complete explanation of I2 I could always refer to D2 as the "proximate" determinant reason of my action of ordering a nonalcoholic beverage. Alternatively, we can say that what "proximately" motivationally guides instrumental rule-following activity is a desire for the ends the relevant rule is instrumental to.

The idea is the following. Human agents are planning agents. Plans come at various degrees of generality or specificity. An individual general plan may contain a series of more specific plans. These more specific plans may form a hierarchical structure.[55] Plan_1 may be part of plan_2, plan_2 may be part of plan_3 and so on. Motivational explanations come at different levels and may resemble the hierarchical structure of plans. So I may say that action A was motivated by some elements in plan_1, plan_1 was motivated by some elements in plan_2, plan_2 by some elements in plan_3 and so on. I suggest that the elements in plan_1 that motivated action A are the "direct" determinant reasons of A, while the elements in plan_2 that motivated plan_1 are the "proximate" determinant reasons of A.[56]

Instrumental rules may be considered elements of plans. If R1 is an element of plan_1, then some elements in plan_2 may have motivated the

54. Desires for the ends that the intentions serve do not exhaust the reasons for having the relevant intentions. There are also reasons for having intentions as an at least satisfying means for achieving the further ends. *See* Thomas Pink, *Purposive Intending*, 100 MIND 343 (1989). Cf. the distinction between "substantive justification" of a rule (which includes reference to the ends a specific rule serves) and "rule-generating justification" (which includes reasons for adopting a rule in order to serve these ends) in FRED SCHAUER, PLAYING BY THE RULES 94 (1991).

55. I follow Bratman in treating plans as mental states rather than as abstract structures. As Bratman contends, plans as mental states may form hierarchical structures; *see* M. BRATMAN, INTENTIONS, PLANS AND PRACTICAL REASON, 28–29 (1999).

56. Elements in the ultimate plan_n of the hierarchical structure provide the "ultimate" determinant reasons for A.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

acceptance of R1. In the discussion above, I single out the desire for the ends the rule is instrumental to. I also argue that R1 cannot be a determinant reason for action. Assuming that A is an act of conformity with R1, we have a choice between two motivational explanations of A. The first is to suppose that the intention to conform to R1, which is an element of $plan_1$, is the direct determinant reason for A. In this case we treat the desire for the ends R1 is instrumental to as an element of $plan_2$ which motivates the acceptance of R1 and as the proximate determinant reason for A. This is in fact ME1. The second is to suppose that the latter desire is also an element of $plan_1$ and the "direct" determinant reason of A. This is ME3.

I will not continue here the task of comparing ME1 with ME3. My aim is to show a) that it is problematic to assume with Shapiro that rules qua rules have motivational clout; and b) that motivational explanations on which rules provide exclusively epistemic guidance are possible and preferable to Shapiro's motivational explanation. The upshot of ME1 and ME3 is that the desire for the ends the rule is instrumental to provides (direct or proximate) motivational guidance to the instrumental rule-follower.

So far I have focused on cases of rule-following in which the preferences of the agent remain the same both at the stage of the acceptance and the stage of the application of the rule. Before concluding, I would like briefly to address Shapiro's account of cases of rule-following in which the agent's preferences are changing over time. In these cases there may be a discrepancy in the preferences of the rule-follower at the stages of rule-acceptance and rule-application. Can my proposed account of rule-following deal with such cases?[57]

Shapiro discusses the case of Barb, who discounts her preferences non-exponentially. Barb has $1,000, which she wants to divide over the year. She budgets on four-month intervals. She does not discount the present at all but she discounts the second four-month period by 33% and the third by 40%. In January when she decides on her budget her maximum allocation is $550 for the first period, $250 for the second period and $200 for the third period. But given that she does not discount the present, in May her preferences have changed. She now does not discount the May-August period and she discounts the September-December period by 33%. Her new preferred allocation in May is $350 for the May–August period and just $100 for the September-December period.[58]

Does Barb's case pose a challenge to my claim that in cases of instrumental rule-following, the desire for the ends the rule is instrumental to provides (direct or proximate) motivational guidance to the instrumental rule-follower? It could be argued that since Barb's preferences have changed, her conformity to her January budget cannot be explained by reference to what motivated her adopting the January budget. There are

57. I am grateful to Scott Shapiro for raising the issue with me.
58. Shapiro, *The Difference That the Rules Make, supra* note 1, at 40–43.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

two relevant questions: the first concerns the rationality of her conformity to the January budget in May; the second concerns the motivation of her conformity in May.

Shapiro's account of Barb's case is incomplete and does not provide us with sufficient information to assess her rationality. First, Shapiro does not specify what caused Barb to adopt her January budget. He says only that the budget is instrumental for her managing her money most efficiently.[59] But this description of the end of her budget does not provide solid grounds for assessing Barb's rationality. How are we meant to judge whether her financial management is efficient? We need to know Barb's long-term plans. It is by reference to them that the efficiency of Barb's budget needs to be assessed. Perhaps some long-term financial plans of Barb render the January budget preferable overall to the May budget. But even if, as presumably Shapiro thinks, this is the case, it does not make any difference to her long-term financial plans whether in May Barb adheres to her January budget or shifts to her May budget, it may be important for some nonfinancial long-term plans that she adheres to her January budget. For example, it may be overall preferable to her that she develops a personality that can commit herself to some decisions even when her short-term preferences change.[60]

Second, apart from focusing exclusively on Barb's short-term preferences, Shapiro's account of Barb's case is incomplete for a further reason. There are different possible cases of conformity with her January budget to which different criteria of rationality apply. Let me single out three such cases: conformity that is accompanied by nonreconsideration of the rule; conformity that is accompanied by recommitment; and conformity that is accompanied by reaffirmation.[61] For example, Barb might have conformed to the January budget in May without reconsidering at all the budget and perhaps without being aware of the change in her short-term preferences. Or she might have conformed after she fully reconsidered the January budget, became fully aware of the change in her short-term preferences, and recommitted herself to the January budget. Finally, she might have conformed after partial reconsideration that fell short of recommitment but nevertheless allowed her to become aware of both the change in her short-term preferences and a reason she had not previously considered in favor of adhering to her January budget. Different criteria of rationality apply to these different cases of conformity. In addition to the rationality of conformity to the budget as such, one should assess the rationality of nonreconsideration, partial

59. *Id.* at 47.

60. Are there cases in which changes in the short-term preferences of the agent have no consequences for any of the agent's long-term plans? I do not want to deny the possibility of this type of Buridan ass cases. But a) it is not clear to me what would be the point of devising and adopting rules for such cases. And b) these cases, though possible, would be rather parasitic cases of instrumental rule-following, the rationality of which is best assessed in cases in which a change in the short-term preferences affects the long-term plans of the agent.

61. For these types of conformity with a rule, I draw freely from BRATMAN, *supra* note 55, chs. 5 and 6.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

reconsideration, and recommitment, respectively. Shapiro's account of Barb's case does not specify Barb's type of conformity, and we are thus unable to assess properly the rationality of her conformity.

Let me now turn to the question about Barb's motivation. If my arguments that peremptory and content-independent reasons are motivationally inert are correct, then Barb cannot be said to conform to her January budget "because it is her budget." So if the January budget does not motivationally guide Barb, what gets her to conform to her January budget in May? Since Shapiro provides incomplete information about Barb's case, I will discuss the three possible cases of conformity that I single out in the previous paragraph. In the first case Barb conforms without reconsidering her January budget and remains unaware of her change in her short-term preferences. In this case what motivates her conformity is her desire for the end for which she adopted the January budget. In the second case Barb reconsiders her January budget and recommits herself to it. She may not recommit herself to the January budget for the reason she adopted this budget in January. For reconsideration, as Bratman argues, "negatives" the link with previous intentions and plans.[62] So we should understand that after she reconsiders her budget in May, she acquires a new intention (that is, different from the one she had in January) to conform to the initial budget that is grounded on her new deliberation in May. Again what gets her to conform to her January budget in May is what gets her to recommit to (or else acquire a new intention to adopt) her January budget.

Finally, in the case of reaffirmation, Barb may become aware of the shift in her short-term preferences. But before opting for a comprehensive reconsideration of her January budget, she may become aware of a reason that she had not previously considered and that supports her adherence to the January budget. For example, she may realize that there is a long-term plan which is advanced by her adherence to the January budget and (before opening the question about the January budget) she might reaffirm her adherence to it. In cases of reaffirmation the agent does not engage in full reconsideration, so neither is the link with her previous intention to conform to the January budget "negatived," nor does she acquire a new intention to adopt the January budget. I suggest that in cases of reaffirmation what gets the agent to conform to a rule is the conjunction of the initial (direct or proximal) motivation for adopting the rule and the new reaffirming reason. So my proposed account of rule-following can be easily modified to accommodate cases of reaffirmation. It need only allow that some reasons, the reaffirming ones, may corroborate the motivational power of the initial reasons for adopting the rule.

To recapitulate, Barb's example gives us no reason to reject our thesis that the desire for the ends the rule is instrumental to provides (direct or proximate) motivational guidance to the instrumental rule-follower. Barb's

62. *Id.* at 94.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press

example unduly mentions only short-term preferences and fails to do justice to the variety of cases of conforming with a rule while one's short-term preferences have changed. We need to take into account the agent's long-term plans and the type of her conformity with the rule to identify the determinant reasons of her conformity.

IV. CONCLUSION

In this paper I have tried to show that Shapiro's attempt to capture the normativity of instrumental rules in motivational terms is unsuccessful. My first set of arguments aimed at showing that the practical-difference thesis fails to do justice to the phenomenon of the determination of action by reasons and thus offers a problematic account of motivational guidance. The second set of arguments tried to show that rules qua content-independent peremptory reasons for action are motivationally inert. Contra Shapiro, rules can provide only epistemic guidance. It is the reasons for which we adopt the instrumental rules that motivationally guide us.

https://doi.org/10.1017/S135232520607011X Published online by Cambridge University Press