

What Personal Rules Can Teach Us About Basic Institutions

CLAIRE FINKELSTEIN*

Suppose I am an alcoholic who wishes to sober up. Assuming I am rational, it should be simple: Because I prefer a sober life to the life of a drunk, I have reason to stop drinking. But, as we know, matters are not that simple. For here I am at a cocktail party. Although I prefer to be dry than to be an alcoholic, I surely prefer to have this one drink than to be dry, at least in the moment. If I am rational, therefore, I will have this drink. But if I reason this way each time, I will always choose to drink, and I will remain an alcoholic.

What is to be done? Perhaps I could check myself into a sanatorium, where I will be physically prevented from drinking, but this has high costs. I would need to take a leave of absence from my job; I would be separated from friends and family. Much better, I think, to quit on my own. So I settle instead on a rule, a rule I will follow in all situations. The rule is this: *do not drink*. Now instead of asking myself whether I would prefer *this* drink to momentary sobriety, I will simply consult my personal rule. And because my rule says *do not drink*, perhaps now I can take a pass at the party.

But the problem is this: If it was rational for me to accept that drink before, how can it be rational for me to follow a rule that says to refrain from drinking? My rule only stands in my way because following the rule means refraining from doing something that will leave me better off, relative to what I now prefer. Having a drink

* Professor of Law and Philosophy, University of Pennsylvania. The Author is the current director of the Institute for Law and Philosophy. She writes on issues at the intersection of moral and political philosophy and the law.

now, after all, need not lead to a further drink; it need not lead to permanent drunkenness. But once again, if I reason this way, I will always choose to drink, and I will end up an alcoholic.

Perhaps therein lies the answer: If I consistently follow the rule, I will be sober. If I do not follow the rule, I will be a drunk. So following the rule will benefit me overall. Thus while not taking *this* drink may be suboptimal, from the standpoint of my current preferences, it is part of a plan of action that is optimal, compared to the feasible alternatives. This, in and of itself, gives me a *reason* to follow the rule.

I have suggested that a person who will be benefited by following a rule has a reason to follow the rule, a reason that makes his following the rule *rational*. Inherent in my suggestion is a more general, and admittedly controversial, claim, namely that an agent has most reason to adopt that mode of reasoning that will benefit him most, even if it commits him to actions that, considered separately, leave him worse off, relative to his preferences. This claim combines what we might call the “benefit principle,” namely that it is rational to make choices according to one’s greatest benefit, with the idea that an agent’s mode of reasoning is a suitable object of choice. Together they suggest that it is rational to select one’s mode of reasoning according to one’s greatest benefit. Thus if following a personal rule conduces to an improvement in one’s overall welfare, as compared with reasoning case-by-case, the benefit principle says that it is rational to reason according to personal rules. Thus it is rational to select actions in accordance with a rule that dictates not drinking—not because it is maximally rational to completely refrain from drinking, but because one can expect to do better by adopting that rule and sticking to it than by following any other rule or by reasoning case-by-case.

Notice, crucially, that there is an important implication of saying that the benefit principle supplies me with a reason to follow a rule, namely that it is not necessary for me to bar myself from reconsidering a personal rule in order to stick to the rule, as other writers in this area have thought. I need not cripple myself mentally, in other words, to stick to a rule on which I have settled. This is even so when I receive new information that might sway me in the direction of abandoning my rule. Suppose, for example, I am standing at the cocktail party, conversing with a cardiologist who is extolling the benefits to me of moderate drinking. I have already settled on my rule of not drinking, and have been successful in sticking to it. The cardiologist, however, tells me that moderate drinking reduces the risk of heart disease, heart attack, and stroke better than intense and regular cardiovascular exercise. This prompts me to reconsider the merits of my rule. While I prefer sobriety to alcoholism, I now prefer moderate drinking to sobriety. Do I now

have a reason to abandon my rule and reach for that drink? And will this not lead me back to case-by-case reasoning, and so to alcoholism? It might be possible for me to revise my rule—to soften it to accommodate my new preferences. I might, for example, adopt a rule like “drink only at cocktail parties,” or “drink exactly one drink a day.” Such rules, if feasible, would be preferable to the more generic rule against drinking. But personal rules of this sort are typically nebulous and easily finessed. They quickly collapse into case-by-case reasoning.

It is for this reason that some believers in personal rules also recommend a policy of nonreconsideration. Reconsideration, they believe, will lead me to side with my preferences and so to abandon my rule. Some reconsideration in the face of new information must be permitted, they allow. But because reconsideration leads to rule abandonment, it must be used sparingly. If the benefit principle is correct, however, I need not side with my immediate preferences when I reconsider. I can side instead with what I take myself to have most reason to do, even in the face of preferences to the contrary. The point is that it is not necessary for me to impair my reflective and deliberative capacities in order to follow a rule.

Notice further that if I did impair my deliberative faculties in order to stick to my rule, I arguably would not actually be *following* the rule at all. Following a rule is a rational activity, distinguishable from various causal methods one might adopt that bypass reason altogether. Thus I might go to a hypnotist or swallow a pill in order to ensure conformity with my not drinking rule. But such methods would be causal, not rational. They would be a form of internalized precommitment—a psychological version of joining a sanatorium. True rule following, however, capitalizes on, rather than eschews, rational agency.

I am interested in the way in which the benefit principle endorses adherence to personal rules, among other things, because I think it is of potential relevance to law. In particular, I believe it provides us with a way of approaching fundamental justificatory questions about basic legal institutions. The question we must ask is whether the rules that govern a given institution are rationally defensible to each person who must live under them, even at moments at which that person would maximize his preferences by abandoning the rule. And this is the question of whether that institution satisfies the benefit principle for each individual contractor whose assent to that institution is required.

The most straightforward application of the benefit principle is in the

area of contract law. It is easy to see that rational contractors are likely to favor an institution that forces them to keep contracts, even in cases in which they would be advantaged by breaking them, as compared with a regime in which contracts were unenforced. The benefit principle, however, will also provide us with more specific guidance for institutional rules. It might make sense of the much maligned doctrine of consideration, since this is a rule that aims to protect mutuality of benefit. And it might explain why it is acceptable for courts to supply a missing term in a contract, provided that they thereby give effect to the parties' actual intentions at the time of contracting, since intentions supply a measure of the benefits each party hoped to receive from the contract.

Can the benefit principle help make sense of legal rules in more far flung settings? For example, can it help justify an institution like punishment or tell us what punishments are legitimate? The question we might ask is whether a rational contractor can regard himself as better off under the rules of that institution than he would be in their absence. To answer *that* question, we must ask whether a person who was himself subject to the rules of the institution would still see himself as faring better, even once he had to incur their costs, than he would fare in their absence. There is reason to think that in many cases, the test will be satisfied. For the deterrent benefits of many penalties will provide individual contractors with greater benefit in the form of security against the infractions of others than they would lose in welfare by being subject to the penalty themselves.

Suppose, for example, we wanted to know whether to institute life sentences for larceny. We would compare the degree of disadvantage to an individual contractor who would suffer that penalty with the degree of advantage to that same individual from living in a society with a rule protecting private property in this way. The latter would be a function of, first, the marginal deterrent benefits of that penalty relative to more moderate penalties, and second, the significance of the substantive value that penalty was protecting. The norms protected by a prohibition on larceny are norms of private ownership, and in the absence of any punishment for larceny (and like crimes), private ownership would be eliminated. Thus each rational contractor should ask himself: Would I be better off under the terms of a contract that established penalties for larceny, assuming that I myself may end up subject to that penalty, than I would be if there were no private ownership at all? Notice that if the penalties for burglary are too low, the deterrent effect will be insignificant, and private property will not be protected. If the penalties are too high, however, agents receiving the penalty would be worse off than they would have been in the absence of private property altogether, and the benefit principle would not be satisfied. The benefit principle, in

combination with the rational goal of deterrence, thus yields specific parameters for the punishment of each separate crime.¹

This last example illustrates a final point about the benefit principle I wish to underline. That principle would not uniquely identify the institutions that rational agents *would* adopt. There are many possible legal regimes that could satisfy the benefit principle. My suggestion is only that rational contracting agents would reject any basic institution that failed the benefit test. Satisfying the benefit principle thus provides a necessary, but not sufficient condition for basic institutions. What additional conditions rational contractors would place on the adoption of basic institutions remains to be developed. In the case of punishment, I suggested the additional condition of deterrent efficacy. Other institutions would have other desiderata. My aim has merely been to show that a certain way of looking at rational agency can give specific and plausible guidance to the task of justifying legal institutions. I shall unfortunately have to leave the details for another day.

¹ For a fuller discussion of this approach to punishment see generally Claire Finkelstein, *A Contractarian Approach to Punishment*, in THE BLACKWELL GUIDE TO THE PHILOSOPHY OF LAW AND LEGAL THEORY 207 (Martin P. Golding & William A. Edmundson eds., 2005).

