Pragmatism and Rationality\(^1\)
Jeremy Randel Koons

In chapter 2, I argued that since moral and epistemic facts do not figure in causal explanations, we must turn to pragmatic reasons to justify continued participation in these practices. What I hope to demonstrate in this chapter is that it is *prudentially rational* to submit to moral and epistemological constraints. Certainly, this has been argued before, at least with regard to moral constraints. However, previous attempts to prove that moral constraints are rational constraints have typically run into various difficulties. In this chapter, I will argue that these difficulties have arisen out of certain systematic misunderstandings regarding the nature of rationality. First, too many philosophers have thought that the individual action is the basic unit of rationality. That is, in questions of means-end rationality, it has too often been thought that we must examine and evaluate individual actions, and determine each isolated action’s rationality in terms of how well that individual action promotes the end in question. I will argue that this is a mistake, and that often strategies, not individual actions, are the basic units of rationality. That is, the rationality of individual actions often cannot be determined in abstraction from the strategy they constitute. Second, philosophers have too often thought of rationality as individualistic. That is, philosophers have thought that we can only evaluate the rationality of actions performed by individuals. I will argue that this, too, is a mistake, and that there is a viable (and indeed indispensable) notion of cooperative rationality.

---

\(^1\) This abridged version of Chapter 3 of Jeremy Randel Koons, *Pragmatic Reasons: A Defense of Morality and Epistemology* (Palgrave Macmillan, 2009) is reproduced with permission of Palgrave Macmillan. This is an extract from the author’s original manuscript and has not been edited. The definitive version of this piece may be found in the aforementioned book, which can be purchased at www.palgrave.com.
Rationality is often ineliminably cooperative, and failure to recognize this fact has led to many puzzles and confusions in philosophy. Let us begin.

Pragmatism and Rule-Worship

So, we have seen that worries about a conflation of practical and theoretical justification are mitigated by a move to a practice conception of rules, which will not allow pragmatic considerations to justify individual beliefs. On our pragmatist account, these practices are structured to promote our interests. On an appropriate understanding of rationality, it is rational to follow the rules constituting these practices.

What do I mean when I say it is rational to follow these practices? I mean that it promotes our interests to do so. Simply put, it is a case of means-end rationality. We have certain interests, and rationality consists of acting to promote these interests. As it turns out, many of these interests are best-promoted by following cooperative strategies. This is what the rest of this chapter aims to demonstrate.

Having defined ‘rationality’ in purely means-end terms, let me pause to assuage some worries. The account here need not deny that there is a notion of constitutive rationality. In other words, I am not denying that we can reason about what our final ends are. If there is some notion of constitutive rationality, then we could use that to establish our rational interests. Morality and epistemology, then, would be strategies aimed at promoting these interests. Thus, even if there is a viable notion of non-instrumental rationality, morality and epistemology are still understood in terms of

---

2 Of course, if this constitutive rationality has a moral element, then one might worry this renders the justification of morality circular. I address this worry in the next chapter.
instrumental rationality: they are strategies aimed at promoting these interests, whether these interests are pre-rational, or constituted by non-instrumental rationality.

An important worry remains to be met, though. It seems as though our two-level strategy (in terms of practices and actions justified by these practices) is vulnerable to an old objection, the rule-worship objection raised by Smart against rule utilitarianism. Rule utilitarianism comes in many different versions, but the basic idea is that we should not test individual actions by the principle of utility; rather, we should test rules (or sets of rules) by this principle. Thus, the set of rules, observance of which (or promotion of which, etc.) would maximize utility is the correct set, and an action is right if it conforms to a correct rule. The charge is that it is often irrational to follow rules, at least when better results can be obtained by ignoring the rule in a particular case.

I am arguing that participation in these practices is rational: that they best promote our interests. The rule-worship objection challenges the idea that it is rational to participate in these practices (i.e., to follow the rules constituting these practices). Smart famously criticized rule utilitarianism as follows.

Suppose that there is a rule R and that in 99% of cases the best possible results are obtained by acting in accordance with R. Then clearly R is a useful rule of thumb; if we have not time or are not impartial enough to assess the consequences of an action it is an extremely good bet that the thing to do is to act in accordance with R. But is it not monstrous to suppose that if we have worked out the consequences and if we have perfect faith in the impartiality of our calculations, and if we know that in this instance to break R will have better results than to keep it, we should nevertheless obey the rule. Is it not to erect R into a sort of idol if we keep it when breaking it will prevent, say, some avoidable misery? Is not this a form of superstitious rule-worship (easily explicable psychologically) and not the rational thought of a philosopher?³

³ Smart (1956), pp. 348-349.
Thus, it might be thought that the practice conception of rules fails, defeated by the Smart’s rule-worship objection. This would be an extremely unfortunate result; indeed, it might even be part of a *reductio* of pragmatism. Consider the consequences of the collapse of the practice conception of rules in a pragmatist system. If this were to happen, pragmatic considerations would be applied directly to individual acts and beliefs. As noted in the previous chapter, epistemology, for example, would no longer be able to serve its purpose. If pragmatic reasons (and not just epistemic reasons) were allowed to justify individual beliefs, then the link between justification and truth would be severed, and our knowledge-seeking practices would cease to be truth-oriented. This would hamstring our pursuit of knowledge—we could no longer take justified beliefs as likely to be true, we could never be sure that our background assumptions and theories were supported by evidence rather than merely pragmatic reasons, and so forth. It would be impossible to pursue any truth- or knowledge-seeking practice in any meaningful way, if we did not keep pragmatic reasons out of our epistemic discourse.

Also, as noted previously, morality would suffer a similar fate. For example, if I promised to do X, then (if pragmatic considerations are applied directly to actions rather than practices) whether I ought to do X is determined wholly by pragmatic considerations, and the fact that I promised to do X has no particular relevance. You can in no way depend upon me to carry through with my promise. Thus, there can be no reliance on promises or contracts, if the pragmatic justification is allowed to trickle down to the level of individual actions.

If rule pragmatism is not viable, and “act pragmatism” (as it might be called) has absurd consequences, then things do not look promising for pragmatism. However, I will
argue that Smart’s objection rests on an over-simplified version of rationality, one that takes individuals and actions as basic. Once we see that in many cases, the basic elements of rationality are teams and strategies, we can defeat Smart’s objection.

In responding to Smart’s objection, we can further elaborate on the idea that it is rational to follow moral and epistemic rules. That is, we can see that instrumental rationality counsels following the rules constituting our moral and epistemic practices. Let us turn now to a demonstration of this important point.

Fundamental vs. Derivative Rationality

The rule-worship charge is that if our ultimate goal is the satisfaction of our interests, then it is irrational to perform an action that ill-serves our interests (even if this action is part of an interest-serving strategy) if there is an available action that would better serve our interests. It is this charge that the pragmatist must defeat.

Human endeavor can be examined at several different levels of detail. We may examine such endeavor at the level of the strategy, or the level of the action (which is, in fact, several different levels; an action can fall under several more or less general descriptions). Actions can be broken down further into sub-actions or even individual bodily movements or parts of movements. The question we must answer is, “If we are to examine the rationality of human endeavor, what level of generality or detail is the proper level for our investigation?”

That this is an appropriate question to ask can be seen by the following example. Suppose I am a maintenance engineer at a hospital where the power has gone out. The best thing for me to do (indeed, the thing I must do) is go to the basement and switch on
the generator. Now, in this situation, is it rational for me to move my right foot 2 inches to the north? The question is virtually without sense. But if this movement is the first part of my first step toward the basement stairs, then the movement makes sense. It makes sense as part of my overall activity, which is going to the basement to turn on the generator. Examined by itself, this movement of my foot cannot be seen as rational. Nor can it be seen as irrational; when viewed in isolation, the rationality of this action cannot be assessed. Rather, it is only rational in the larger context of my efforts aimed at restoring power to the hospital. In other words, moving my right foot 2 inches north might be rational (although this claim has, to my ear, an awkwardness to it), but its rationality is wholly derived from the rationality of my more general action: going to the basement to turn on the generator.

We can therefore distinguish between derivative rationality and what we might call fundamental rationality. Moving my foot 2 inches to the north is derivatively rational. If that were all I were to do (if I stopped moving my foot after 2 inches), this action would not be rational; the action is only rational because it is part of the overall action of going to the basement to turn on the generator. This more general action has what one might call fundamental rationality—it is the rational action, in the circumstances, and all sub-actions and movements (such as moving my foot 2 inches to the north) inherit or derive their rationality from the rationality of this more general rational action.

Thus, finding the right level of description is fundamental to rationality. This is no more surprising than the claim that we cannot understand biological speciation in terms of quarks; it is merely a recognition that if a subject matter is comprehensible at
one level of description, it is not necessarily more comprehensible (or comprehensible at all) at a more detailed level of description. And the problem is not merely epistemological—it is not merely one of comprehensibility. Arguably, there simply are no facts about biological speciation that can be specified at the level of quarks. Similarly, there simply won’t be many facts about fundamental rationality if we focus on movements at the level of “moving ones foot 2 inches to the north.” As I argued above, this action taken by itself simply is not rational (and perhaps not even subject to rational evaluation). Its rationality can only be understood derivatively, as parasitic on the fundamental rationality of a more general action, such as going to the basement to turn on the generator.

So what is the correct level of description for the study of rationality? The answer is that it depends on the particular circumstances of action. We can say, though, that the rule worship objection rests on the dogma that the level of the individual action is always the correct level on which to understand rationality. This is a dogma equivalent to the positivist dogma that the level of physics is the only correct level at which science can be understood. Just as the positivists were mistaken (chemistry, biology, and astronomy are real sciences even if they are not reducible to physics), Smart and company are equally mistaken in assuming that rationality must be understood exclusively at the level of individual actions. It is to this argument that we will now turn.

**Actions and Strategies**

The fundamental unit of rationality is often the individual action. If I am standing before the generator switch in a hospital that has lost its electricity, the rational thing to
do is to flip the switch. It seems clear that this action is fundamentally rational—it need not possess merely derivative rationality in virtue of being part of some larger action. But there are cases where the smallest unit of fundamental rationality is not the individual action, but the strategy. We can demonstrate this with the use of a simple example. Suppose I am on a diet. If I break down and have a scoop of ice cream, I will get a great deal of pleasure from this ice cream, and the amount of weight I would gain (a fraction of an ounce, no doubt) would be utterly unnoticeable to me. So the act of eating a scoop of ice cream has overall positive consequences. But if every day I thus weigh the consequences of eating a scoop of ice cream, and decide to do so, then in no time I will be overweight, with all of the unhappy consequences that this will bring. So it seems clear that if my goal is my long-term happiness, I cannot merely look at individual actions: no individual instance of eating a scoop of ice cream has any long-term consequences (good or bad); it has only positive short-term consequences. Instead, I must think in terms of strategies. Failing to do so will result in my long-term unhappiness. Hence, rationality—in this case, the achievement of my long-term happiness and good health—demands that I think at the level of strategies, not individual actions. Treating the action as the unit of fundamental rationality here ultimately leads to the frustration of my long-term goals. Thus, treating the action as the unit of fundamental rationality is ultimately self-defeating. We must instead realize that in this case, my goals can only be met by focusing on the level of the strategy. Thus, in this case, the strategy is the proper level to focus on, if we want to best achieve our goals.

Although my case is hardly weakened if this statement is wrong. The less individual actions are bearers of fundamental rationality, the stronger is my argument against the rule-worship objection.
This is a case of means-end rationality. Failure to achieve the desired end due to poor action selection signals a failure of rationality. But if I treat individual actions as the bearers of fundamental rationality, then I will fail to achieve my desired end, and I will fail because I selected the actions I did. This signals a failure of rationality. And so we must reject the idea that in this case, the individual action is the fundamental unit of rationality. I will only succeed in achieving my end if I pursue the overall strategy. On each occasion, passing up the ice cream is the rational thing to do—but this rationality is derivative; it is parasitic on the rationality of the overall strategy, just as the rationality of moving my foot 2 inches to the north was parasitic on the rationality of going to the basement and turning on the hospital generator. In both cases, examining the issue at the wrong level of detail leads to the wrong answer. If I examine moving my foot 2 inches to the north, I can only conclude that this is not a rational response to a power failure. But it is rational, in the larger context of restoring power to the hospital. Similarly, refusing a scoop of ice cream on a particular occasion is not rational, if viewed in isolation. But viewed in the context of my goal to lose weight, then it is rational—but its rationality derives from the rationality of the overall strategy of refusing ice cream (or perhaps having ice cream only once per week, or whatever the optimal strategy turns out to be, balancing the pleasure of eating ice cream against my desire to lose weight). Thus, if what we care about is achieving our interests, and rationality is defined instrumentally in terms of achieving our interests, then it is irrational to reason at the level of actions rather than strategies, for doing so will fail to promote our interests. Rationality demands thinking at the level of strategies.
The act utilitarian might object as follows. Suppose I eat 100 ice cream cones over 100 days, and as a consequence have gained 10 pounds at the end of this 100 days. As a result of my weight gain, I am unhappy. Let us suppose that although I gained 100 utiles from eating the ice cream cones, this increased weight gain has caused me -110 utiles. Thus, there is a net loss of 10 utiles. My anti-act utilitarian argument above goes as follows: “Each day, I gained 1 utile from eating the ice cream cone. I was unable to notice any incremental weight gain, and so didn’t feel any suffering. Thus, on any given day, there was an act utilitarian reason for eating an ice cream cone. But at the end of 100 days, I actually suffered a net loss of 10 utiles. This shows that if I really care about promoting my happiness, I ought not be an act utilitarian.”

The act utilitarian will respond that the problem is with my accounting, not with act utilitarianism itself. If I do, in fact, get -110 utiles from eating all of these cones, and this loss of utility really is the result of eating all of these ice cream cones, then the sensible thing to do is to divide up that negative utility and apply it to each cone. Thus, I may gain 1 utility of pleasure from eating the cone, but I also lose 1.1 utiles from the unnoticeable (but nevertheless very real) incremental weight gain caused by eating this cone. I may not notice the weight gain, but the act utilitarian is, of course, interested in long-term utility. Thus, the act utilitarian thing to do is to have an ice cream cone only occasionally.

This act utilitarian response will not work. It assumes that a gain of a fraction of an ounce in weight is in itself a bad thing. The act utilitarian assumption is that (if one is within a certain weight range) any gain of weight (even 1/4 ounce) must count as a bad
thing, and any loss of weight (even \(\frac{1}{4}\) ounce) must count as a good thing. However, a person’s weight naturally fluctuates by ounces or even pounds. Should we treat this natural fluctuation as a series of good and bad events? The loss and subsequent regaining, over the course of a 24-hour period, of \(\frac{1}{4}\) of an ounce strikes me as a matter of indifference. It does not strike me as a good event followed by a bad event. Slight changes in a person’s weight simply do not seem to be good or bad events; they are indifferent. A gain of \(\frac{1}{4}\) ounce does not affect a person’s health, or mobility, or flourishing in any substantive way. So it is not obvious why the gain of a fraction of an ounce of weight deserves to count as a loss of 1.1 utile.

Of course, the act utilitarian can respond, “Granted, the gain of \(\frac{1}{4}\) ounce is not by itself significant. But in the original example, we were talking about the gain of several pounds over 100 days. This results in significant disutility, and so it makes sense to assign a fraction \((1/100)^{th}\) of the total disutility to each ice cream cone consumed.” This response may be correct, but notice that this response is out of step with the basic atomistic assumptions of act utilitarianism. The response concedes that one ice cream cone by itself would not create any disutility, but that considered in aggregate, the consumption of 100 cones creates a great deal of disutility, and this disutility can reasonably be divided among the individual actions. Notice, however, that the act utilitarian must think first in terms of groups or sets of actions; the disutility of the set of actions is prior to the assigning of disutility to the individual actions in the set. That is, the disutility of the individual cone derives from the fact that the person eats 100 cones, not just 1 cone. Thus, the disutility of the set of actions is prior to the disutility of the individual action. If the individual action existed in isolation—if the person ate only one
Pragmatic Reasons—Chapter 3

cone, not 100—then there would be no disutility from this individual action (or so I have been arguing). So the act utilitarian is already being forced to think like a rule utilitarian—she is being forced to think about the consequences of sets of actions (dare I say plans or strategies?) before she can calculate the utility of any individual action constituting this set. The utility of the action derives from the utility (or disutility) of the set of actions.

But of course, this is the line I have been pushing all along—the rationality of an individual action may, in many cases, be derivative from the rationality of a strategy or a larger set of actions. So the above act utilitarian response actually makes a crucial concession—it concedes that the rationality of individual actions is not necessarily fundamental—that it can instead be parasitic on the rationality of a long-term plan or set of actions. Eating one ice cream cone creates no disutility. But if I plan (or foresee) that this ice cream cone is the first of 100, to be consumed over 100 days, then I must treat this action as having a net utility of -0.1. The rationality of the action derives from the rationality of the plan or strategy.

Let us continue with the thread of our discussion. When talking about justifying a normative practice, a number of problems can be solved by moving from a picture where individual acts are justified by appeal to our interests to a picture where strategies are so justified, and acts are seen as rational or justified insofar as they relate to these strategies. Consider some hoary examples from the utilitarian literature: water conservation during a drought and taking a shortcut across the grass in a well-traveled area. In both cases, ill effects will result only if a certain number, n, of people fail to conserve water or take the shortcut across the grass. Now, it is a well-known objection against act-consequentialism...
that each person ought to reason as follows: “My violation of the water conservation order (or my taking a shortcut across the grass) will have no appreciable effect on whether the city suffers a water shortage (or the grass dies), but will cause me a measurable amount of utility. Therefore, it is utility maximizing for me to violate the water conservation order (or walk across the grass).” Of course, since it is rational for each person to think this way, then if everyone is an act consequentialist, all will violate the water conservation order (or take the shortcut across the grass), and hence the city will suffer a water shortage (or the grass will die). Thus (it is argued) the best consequences can only be achieved if people don’t behave like act utilitarians, but instead follow a rule: “Conserve water” or “Don’t walk on the grass.”

Act utilitarian responses to this objection occasionally involve using mixed strategies. Smart, for example, says that an act utilitarian ought to find out how many people can violate the water conservation order and then employ a mixed strategy to determine whether or not she ought to comply or violate. Thus, suppose that 10% of the city’s residents can violate the water conservation order without any serious negative consequences resulting. Smart recommends, then, that the act utilitarian use a random procedure to determine whether or not she should violate, a procedure that will with a probability of .9 recommend that she not violate the water conservation order. Similarly, if n people (out of a total number m of people who will have the choice of walking across the grass or not) can cross the grass without it dying, then the act utilitarian might suggest that people employ a mixed strategy with probability of n/m of allowing each person to walk across the grass; or the utilitarian might suggest that the n oldest or most mobility-impaired people be allowed to take the shortcut.
There are two comments to be made about these act utilitarian replies. The first, and most obvious, reply is that these responses are not, in fact, act utilitarian responses. All of them counsel following a strategy. Consider Smart’s recommendation that we employ a mixed strategy. This is not an act utilitarian move. For an individual act utilitarian will reason as follows: “My defecting from this mixed strategy and simply deciding outright not to follow the water conservation order will have no appreciable effects on the city’s attempt to conserve water. However, defecting has a higher expected utility than following the mixed strategy. If my obeying the conservation order will create 0 units of utility, whereas my violating it will create 1 unit of utility, then following the mixed strategy has an expected utility of .1, whereas violating has an expected utility of 1. Therefore, I should violate.” Following a mixed strategy is just that: it is following a strategy. As such, it is a concession that utilitarianism can only get the job done if strategies are employed; as we just saw, a strict act utilitarian will not follow the strategy.

The second comment that needs to be made about these act utilitarian responses is as follows: a moral system must be applicable. That is, its elements must be learnable, and it must be possible for real people to follow and apply these elements. As Daniel Hunter writes, “There are cognitive constraints on the acceptance of rules. The rules, for example, cannot be too complicated, else people simply couldn’t remember them.” Rawls makes a similar point, writing,

\[
\text{[P]rinciples are to be universal in application…Thus I assume that each can understand these principles and use them in his deliberations. This imposes an upper bound of sorts on how complex they can be, and on the kinds and number of distinctions they draw.}\]

\[\footnote{\text{Hunter (1994), p. 32.}}\]

\[\footnote{\text{Rawls (1971), p. 132.}}\]
However, it is sheer fantasy to suppose that the act utilitarian “solutions” suggested above are applicable. How is the average person to know how many people can violate the water conservation order without serious consequences resulting? How is the average person to know how many people can cross the grass before the grass dies? Can the average person really do a good job of calculating probabilities (including conditional probabilities, which would be required in more complex cases) and employing a suitable mixed strategy? And utilizing these strategies often assumes that people will, on short notice, all be able to come up with suitable strategies (such as allowing the n most infirm people to cross the grass). Again, this is fantasy. It is not a livable moral system. Thus, we can see that in a rule utilitarian system, strategies produce the best consequences; and so a system that treats strategies as fundamentally rationally will produce better consequences than a system that treats actions as in all cases the bearers of fundamental rationality. Consequentialists ought then to use considerations of utility to justify strategies, not individual actions.

Let us consider another argument for the conclusion that rationality need not be understood in terms of individual actions, an argument advanced by Richard Dean.9 The dogma that rationality must be understood in terms of individual actions is equivalent to embracing the theory of straightforward maximization (or “s-max”, as Dean abbreviates it). According to s-max, “making a rational choice is simply a matter of choosing the action that provides you with the maximum expected utility.”10

---

9 Dean (1997).
David Gauthier famously defends the rational theory of constrained maximization.¹¹ Dean writes,

Constrained maximizers usually seek to maximize their own expected utility, but in specific sorts of cases are willing to forego a maximizing action in favor of a cooperative one. In a PD [prisoner’s dilemma], constrained maximizers will act cooperatively provided they think others will cooperate as well. For the sake of engaging in mutually beneficial projects with others, they will forego some opportunities to maximize their utility.¹²

Gauthier argues that constrained maximizers will achieve more utility in the long run, since cooperative opportunities will be open to them that would not be open to the straightforward maximizer. The problem of pursuing strategies, as I have outlined it here, is similar to Gauthier’s problem of cooperation: both Gauthier and I are trying to argue that it is rational to perform a non-maximizing action, if that action is part of a policy or strategy that will maximize utility in the long-run. How can we show that it is rational to perform a non-maximizing action?

Dean argues that the key to solving this puzzle is to distinguish between the aim and policy components of a theory.¹³ The aim component of a theory is essentially a theory of value. It “defines the objects that have value for agents.”¹⁴ The policy

¹¹ Gauthier (1986).
¹³ Dean (1997), p. 457. This distinction is similar to Bales’ (1971) distinction between an ethical theory’s account of right-making characteristics of actions and the theory’s decision-making procedure. I will not discuss Bales’ article, as I hold it to be fatally flawed. Bales argues that act utilitarianism could still be a correct account of the right-making characteristics of an action, even if it is not correct as an ethical decision-making procedure. But if act utilitarianism is correct as an account of right-making characteristics, then it seems as though a correct ethical decision-making procedure must (in an individual case) lead us to the utility-maximizing action. But this focus on individual actions is, as I have argued, self-defeating. If utility is what we ought to maximize, we ought not be prejudiced at the outset in favor of act or rule utilitarianism as an account of the right-making characteristics of actions. It may turn out that the utility-maximizing action is not the morally correct one (the focus here being on right-making characteristics)—if the action is forbidden by a utility-maximizing strategy. And I have been arguing for a conclusion similar to this one.
Now, the standard theory of rationality says that a rational agent’s aim should be maximization of her expected utility. However, the standard theory then hastily concludes that if this is the agent’s aim, then each action should be aimed directly at this end. In other words, there is a (perhaps conceptual) link between the aim and policy components of a theory: if the aim is X, then every action ought to be aimed directly promoting X.

But why should we make this hasty leap from aim to policy? Dean writes, “Rather than assuming that this aim must be accompanied by a policy of directly seeking the aim, however, Gauthier maintains that the correct policy is whatever policy best fulfills the aim.” And of course, this makes perfect sense. Given a theory of value, which defines the objects worthy of pursuit, it is rational to pursue whatever policy best promotes this aim, and irrational to adopt a policy which you know will fail to achieve this aim as well as another available policy would. This seems obvious. But it leads to the interesting conclusion that if maximization of personal utility is one’s aim, then the policy of straightforward maximization is not the rational one to pursue. Rather, a policy of designing and executing strategies will often be dictated by the policy element of the theory.

Similarly, if our aim is to best promote our interests, a policy of choosing actions in isolation will not be the best one. Rather, we should often pursue strategies. Even though these strategies will often recommend actions that are contrary to our interests, these strategies are rational insofar as they are the best available policy for promoting our

---

interests. The assumption that straightforward maximization must always be rational is a dogma, and cannot withstand critical scrutiny. We should often reason in terms of strategies and rules, not in terms of individual actions.

Thus, we must construe our pragmatism as rule pragmatism. If we construe it as act-pragmatism, then our pragmatically-justified practices will cease to achieve their ends. Consider our previous observations about epistemology: as I argued above, if pragmatic reasons (and not just epistemic reasons) were allowed to justify individual beliefs, then the link between justification and truth would be severed, and our knowledge-seeking practices would cease to be truth-oriented. This would hamstring our pursuit of knowledge—we could not longer take justified beliefs as likely to be true, we could never be sure that our background assumptions and theories were supported by evidence rather than merely pragmatic reasons, and so forth. It would be impossible to pursue any truth- or knowledge-seeking practice in any meaningful way, if we did not keep pragmatic reasons out of our epistemic discourse. And so we must treat strategies as fundamental, and adhere to our practice conception of rules.

**Individual Strategies and Team Strategies**

We have seen that in many cases, the strategy (not the individual action) is the basic unit of fundamental rationality. Thus, the rule-worship objection goes astray in accusing someone of irrationality if she insists on following a strategy even when the strategy dictates sub-optimal actions.

However, if we look at the above examples, we note that the problem facing the act utilitarian is not merely one of failing to follow a strategy. The problem is also one of
cooperation. For the problems that occur if one thinks in terms of *acts* rather than *strategies* recur of one thinks *individualistically* instead of *cooperatively*.

Consider the matter this way. Suppose I, as a pragmatist, am convinced by the above considerations that I should always think in terms of strategies, rather than in terms of individual actions. Does this mean that I will therefore always cooperate (i.e., conserve water, not walk on the grass, etc.)? Not necessarily. I might reason as follows: “A strategy of non-cooperation will best-serve my interests. For my non-cooperation will not result in significant negative consequences, but will in fact benefit me. Thus, overall consequences are maximized.” Thus, the pragmatist might follow a strategy of *free riding*: she recognizes that her failure to cooperate is insignificant in large-scale cooperative endeavors (such as water conservation and grass management), and so the best individual strategy is to free ride. Of course, if everyone reasons this way, then all follow a strategy of free riding, and we are deprived of the obvious benefits of cooperation. What is to be done? The solution is to recognize that not only is the strategy often the smallest unit of fundamental rationality, but also that rationality is often *irreducibly cooperative*.

The idea that there is a notion of team rationality, not reducible to individual rationality, is advanced by Margaret Gilbert\(^\text{17}\), Robert Sugden\(^\text{18}\), and others. As with the ice cream example above, where we showed that it is often irrational to think in terms of actions rather than strategies, we can show that rationality is often cooperative if we can give an example where it is clear that rationality can only be construed cooperatively, not

---

17 See, for example, Gilbert (1989) and (2001).
18 Sugden (2000).
individually. Sugden presents such an example with what he calls the Footballers’ Problem.

Suppose that A and B are two attacking players in a football team. A has the ball, but a defender is converging on him. B has more space, so A wants to pass the ball to him. There are two directions in which B could run so as to intercept a pass from A: call these left and right. Correspondingly, there are two points on the field, left and right, to which A could pass the ball to be picked up by B. There is no time for communication, or for one player to wait to see what the other does: each must simultaneously choose left or right. Suppose that the move to the right puts B into a slightly better position. Say that the probability that the pass will result in a goal is 10 per cent if both choose left and 11 per cent if both choose right. If one chooses right and the other left, the probability is zero. What should each player do?19

The Footballers’ Problem can be presented in a traditional decision matrix as in table 3.1:

<table>
<thead>
<tr>
<th></th>
<th>player B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>left</td>
<td>right</td>
</tr>
<tr>
<td>player A</td>
<td>10,10</td>
<td>0,0</td>
</tr>
<tr>
<td></td>
<td>0,0</td>
<td>11,11</td>
</tr>
</tbody>
</table>

What should the two players do? “The answer seems obvious: each should choose right,” writes Sugden. “But paradoxically,” he continues, “this obvious answer cannot be generated by the theory of individual rationality, as used in game theory.”20

Sugden explains as follows:

Clearly, if A expects B to choose ‘right’, he should choose ‘right’ too. But equally, if A expects B to choose ‘left’, he should choose ‘left’ too. According to expected utility theory, A should choose ‘right’ if he judges

the probability that B will play ‘right’ to be greater than 10/21; and he should choose ‘left’ if he judges that probability to be less than 10/21. So what is rational for A to do depends on what B can be expected to do. In situations like this, game theory invokes the assumption that the rationality of the players is common knowledge between them. Thus, in order to form a rational belief about what B will do, A has to take account of the fact that B is himself rational, and so will choose whatever is rational from his point of view. But B’s problem is exactly symmetrical with A’s: it is rational for B to choose ‘right’ if A can be expected to play ‘right’ with a probability of 10/21 or more, but rational for B to choose ‘left’ if A can be expected to play ‘right’ with a probability of less than 10/21. We have entered an infinite regress: what it is rational for a player in a situation like A’s to do depends on what it is rational for a player in a situation like A’s to do.\(^\text{21}\)

The solution to the problem is for A and B to think not about what is rational for each of them to do, but what is rational for the team to do. And it is clearly rational for the team (meaning A and B) to go right. Again, though, this solution cannot be reached from the perspective of individual rationality; it can only be a product of team rationality. And so we have an example where rationality is irreducibly cooperative: it is a question of what we ought to do, and cannot be reduced to separate questions about what I ought to do and what you ought to do.\(^\text{22}\)

One might argue that the Footballers’ problem is misleading, as the two players both have the same goal. But cooperative rationality is often called for even if the players have different goals. Consider the venerable prisoner’s dilemma. I will list the outcomes for each player in order of preference in table 3.2:


\(^{22}\) As Sugden points out, it will not help to try to understand this problem from the point of individual rationality by looking for a Nash equilibrium. Not only are there three different Nash equilibria in this problem, but more fundamentally, “the fact that a particular combination of strategies is a Nash equilibrium (even a unique Nash equilibrium) gives neither player a reason to choose to play his part in that equilibrium. All we can say is that, if it were the case that each player expected every other player to play his part in a certain (strict) Nash equilibrium, then each player would have reason to do the same. And that provides us with no escape route from the infinite regress of reasons” (p. 182).
Table 3.2

<table>
<thead>
<tr>
<th></th>
<th>player B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cooperate</td>
<td>defect</td>
</tr>
<tr>
<td>player A</td>
<td></td>
<td>2,2</td>
</tr>
<tr>
<td></td>
<td>cooperate</td>
<td>1,4</td>
</tr>
<tr>
<td></td>
<td>defect</td>
<td>4,1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3,3</td>
</tr>
</tbody>
</table>

The puzzle of the prisoner’s dilemma is that it is rational for each person to defect, though if each person defects, then each is worse-off then he or she would have been had they cooperated. Notice, however, that this understanding of the prisoner’s dilemma essentially involves individual rationality. If we ask, for each player, what it is rational for that player to do, the correct answer is “defect.” But this leads to the paradoxical result that each player is not very well off, achieving only his/her third preference (out of four). But if we instead ask what it is rational for the players (taking as a collective) to do, then the answer is “cooperate”: if the players cooperate, then they will both achieve their second preference. But we can only achieve this happy result if we ask what it is rational for us to do; if we insist on asking what it is rational for each player, considered individually, to do, then the players will continue to achieve a sub-optimal result. Individual rationality is no replacement for cooperative rationality.

I have been arguing that if a particular style or method of means-end reasoning systematically fails to achieve the desired end, where another style of reasoning would achieve that end, then it is irrational to follow the first style or method. And since in the prisoner’s dilemma individual reasoning leads to the next-to-last (third) preference, whereas cooperative reasoning leads to the second-best (second) preference, individual reasoning fails the minimal test of means-end reasoning: it systematically fails to achieve
the desired goal, where another style of reasoning reaches this goal. Thus, the prisoner’s
dilemma again shows that (a) cooperative rationality is not reducible to individual
rationality, and (b) cooperative rationality is often superior to individual rationality.

Thus, rationality must often be understood cooperatively. Again, if what we care
about is achieving our interests, and rationality is defined instrumentally in terms of
achieving our interests, then it is often irrational to reason individually rather than
coopertatively, for doing so will fail to promote our interests. Rationality demands
cooperative strategies.

Act consequentialists have made numerous attempts to sneak cooperation into
their account, while clinging to a model of act-based individualistic rationality (as
opposed to strategy-based team rationality). Brandt, for example, has argued that
defection (e.g., breaking a water conservation order) is irrational for the following reason.
Suppose it would maximize utility for you to break the water conservation order. If it is
rational for you to do so, then it is rational for all others to do so as well. But since all
others doing the same would result in overwhelmingly negative consequences, it is
therefore not rational for you to break the water conservation order. However, Brandt is
trying to sneak team rationality in through the back door. If we insist on individual
rationality, then we can only ask what it is rational for you to do. And if your breaking
the conservation order would not cause others to do the same, then it is utility-
maximizing for you to break the order, and irrational for you not to do so. But if we
move to a model of team rationality, we can see that it is rational for us to obey the water
conservation order, and hence to restrict our water usage.
Many will be suspicious of the idea of cooperative rationality. This suspicion can be brought out more clearly by focusing on differences between the Footballer’s Problem and the ordinary prisoner’s dilemma. In the Footballer’s Problem, both players have the same goal—namely, to score a goal. Thus, cooperation best serves the goals of each player. It is rational for (e.g.) player A to cooperate because by cooperating, player A maximizes his personal utility.

However, the key feature of the Prisoner’s Dilemma is that cooperation does not maximize one’s personal utility—defection does. This is based on the assumption that the player’s strategies are independent—that is, the strategy chosen by the other players does not depend on the strategy you choose. In this case, the critic of cooperative rationality will argue that cooperation is irrational, since it fails to maximize one’s personal utility.

The first thing to say in response to this charge is that this objection again assumes that rationality can only be understood individually. But that is the very dogma we are rejecting here. It may be true that my cooperation does not maximize my personal benefit—but the point is that our cooperation maximizes our benefit. By cooperating, we realize our shared goals, and so it is rational for us to cooperate. To assume that all questions of rationality can be answered in terms of how it is rational for me, considered individually, to do X, is to beg the question against the current account. The Footballer’s Problem, though not a typical prisoner’s dilemma, has already shown that questions of what it is rational to do cannot always be given an individualistic answer.

Let us approach the matter from another direction. An incautious thinker might conflate satisfaction of my preferences with maximization of my well-being. But this
conflation assumes that all of my preferences or interests are self-regarding. That is to say, this conflation tacitly assumes the truth of psychological egoism, the idea that all of my preferences and desires are directed toward producing the maximum possible benefit for myself. If this were true, then the realization of my goals would be independent of the realization of your goals. If my interests and your interests are all self-regarding, then realization of my interests is at best independent of realization of your interests, and at worse conflicts with the realization of your interests (as when scarce resources are at stake). Even if a game theorist doesn’t make this egoistic assumption, the prisoner’s dilemma is still geared toward individualistic rationality: it lists my utilities and your utilities, but it lists them separately, and cannot convey the degree to which they overlap and reflect shared goals and preferences.

However, this assumption regarding the independence of our interests cannot be made with respect to other-regarding interests. If we reject the assumption of psychological egoism, we clear the way to a richer understanding of rationality. It is true that we have self-regarding interests. Indeed, our interests are often self-regarding. But humans are social animals, and as such have other-regarding interests. Thus, consider A and B, who are members of the same community. A wants A to flourish, of course, but he also wants the members of the community to flourish. B also wants B to flourish, but he, too, wants the other members of the community to flourish. So if we focus separately on A’s preference satisfaction and B’s preference satisfaction (as traditional game theory asks us to do), we ignore the fact that A and B share the same interest—namely, the interest in the flourishing of their community. Since the members of the community share this interest, and since coordination among the members of the community will
greatly increase their success in promoting this interest, it is rational for the members of the community to coordinate in order to promote this shared interest.

Thus, egoistic assumptions might lead one to treat our preferences as separate and distinct. But if we reject this assumption, we can see how the members of the community essentially share certain interests. Given the existence of community-wide interests, it makes it easier to see the rationality of a community-wide cooperative strategy. The existence of shared interests, then, is one way to argue that cooperative rationality is a legitimate notion. We will return to this idea shortly.

But shared interests are not the only realm in which cooperative rationality is a useful idea. A cooperative moral strategy will also advance my self-regarding interests. Gauthier writes that a moral system has several attributes: first, “Morality is a system of principles such that it is advantageous for everyone if everyone accepts and acts on it.” Second, “acting on the system of principles requires that some persons perform disadvantageous acts.” Third, “each person must gain more from the disadvantageous acts performed by others than he loses from the disadvantageous acts performed by himself.” Fourth, “Each person will do better [if a moral system is adopted] than if no system is adopted.” In essence, then, although morality requires that I sometimes act contrary to my self-interest, I benefit from the existence of a cooperative moral arrangement, and am better off than I would be in the absence of such an arrangement. As this is true for everyone in the community, it follows that each of us is worse-off if there is no such cooperative strategy in place. Thus, our self-interest is frustrated if the

---

23 Gauthier (1967), pp. 461-2, emphasis removed from original. Gauthier attributes this position to Kurt Baier; it is not clear that Gauthier himself fully endorses this claim.
24 Gauthier (1967), p. 462, emphasis removed from original.
community fails to adopt such a cooperative strategy. Since failure to achieve ones ends through adoption of a bad strategy (or failure to adopt a particular strategy) signals a failure of means-end rationality, it follows that it is not rational for a community to refuse to adopt such a strategy. It essentially frustrates each person’s ends.

Notice, however, that only the supporter of cooperative rationality can explain why such a failure is irrational. The supporter of individual rationality admits that each person is worse-off, if he thinks individualistically; but the supporter of individual rationality is forced to chalk this up as a paradox of rationality. Her view does not make room for an explanation of why failure to cooperate is irrational. That is to say, although the choices made by the members of the community thereby preclude these members from achieving their desired goals (maximization of advantage), the supporter of individual rationality cannot explain why such failure signals a failure of rationality—for the simple reason that for this supporter, no failure of rationality has occurred—even though the community members’ choices prevented them from achieving their desired ends. If we want to achieve our desired goal—that each of us is as well-off as possible—then it is rational for us to follow a cooperative moral strategy.

The argument here is in some ways related to the argument, noted earlier, advanced by Brandt. You will recall that Brandt advanced the following argument against defection from morality. Suppose it would maximize utility for you to defect. If it is rational for you to do so, then it is rational for all others to do so as well. But since all others doing the same would result in overwhelmingly negative consequences, it is therefore not rational for you to defect. What prevented Brandt’s argument from being sound is that Brandt was working with a notion of individual rationality. If we ask what

---

it is rational for me to do, and consider that my actions will have only minimal (or no) effect on the actions chosen by others, then we can only conclude that it is rational for me to defect. But if we allow the notion of cooperative rationality, then the fact that all of us doing X would harm each of us does provide an argument for us not to do X. But this argument can only establish that it is rational for us to refrain from X—that is, it essentially depends on the notion of cooperative rationality. Again, if we confine ourselves to the question of what it is rational for me to do, and if my defection would not cause others to do the same, we can only conclude that I ought to defect, and it would be irrational for me cooperate.

If we combine (a) the existence of other-regarding interests with (b) the notion that only by following a cooperative moral strategy can we best-advance our self-regarding interests, we can see that the notion of cooperative rationality not only makes sense, but is a powerful and viable substitute for individual rationality in many contexts. In other words, a rejection of psychological egoism, coupled with a recognition that rationality is often both strategic and cooperative, is essential to understanding institutions such as morality.

I will make one final comment. As noted above (in the section “Actions and Strategies”) Gauthier’s argument that cooperation is often rational. However, I will not pursue Gauthier’s argument here. The reason is as follows: Gauthier is trying to prove from the standpoint of individual rationality that it is rational for the individual to cooperate. Thus, Gauthier is still taking the notion of individual rationality to be basic. But what I am trying to establish here is that there is another notion of rationality, cooperative rationality, which is separate from and not reducible to individual rationality.
A demonstration that it is individually rational to cooperate with others does not show this. It is not an unwelcome addition to the theory—of course, if we want to show that it is rational to be moral, then it is helpful to be able to show that it is both individually and cooperatively rational to be moral—but the goal of the current section is to show that the notion of cooperative rationality is viable in the first place. Showing that it is individually rational to cooperate does not show this, but merely relies on the pre-existing notion of individual rationality. However, I will note that to the extent that Gauthier and others have demonstrated that cooperation is individually rational, they have strengthened my case that it is rational to be moral. Overdetermination never hurts anything.

To conclude this section: the general pragmatist strategy is to show that it is rational to be moral. Morality consists of a set of strategies aimed at promoting our interests. Discussions of whether it is rational to be moral in general discuss whether it is always prudent to be moral, where ‘prudence’ is treated as synonymous with ‘self-interest.’ But we must recognize that in addition to self-regarding interests, we also have other-regarding (i.e., altruistic) interests. I will have much more to say about this in the next chapter, when I discuss the pragmatist account of morality. But for the time being, let us note that the upshot of this section is that following morality displays means-end rationality, where the end in question is the promotion of our interests.

**Collective Preference and Joint Intentions**

What grounds collective action? What makes it rational to participate in a cooperative strategy? I think that in acknowledging the viability and legitimacy of
cooperative rationality, we should also recognize the existence of something like group preferences, joint commitment and/or joint intentions.

*[the remainder of the section has been omitted from the on-line sample]*

**Further Objections against Rule-Based Systems**

The pragmatist system I am advancing here is in some sense teleological, in the sense that the rules are aimed at promoting some good—namely, they are aimed at promoting our interests. As such, the theory is vulnerable to the objections that are raised against rule utilitarianism, that other well-known rule-based teleological system. Let us examine some of the better-known objections to see if they carry any weight.

**The ‘Simple Collapse’ Argument**

One objection against rule utilitarianism is that it collapses into act utilitarianism. A similar charge might be leveled against rule pragmatism. Suppose you have a practice, consisting of a set of rules. Consider a rule, R1: “Whenever in circumstances C, do action A.” However, the objection continues, when in C, the agent is actually faced with a choice between two different rules:

R1: Whenever in circumstances C, do action A.

or:

R2: Whenever in circumstances C, do action A, or whatever action will produce the best results.

Clearly (the objection goes) the second rule will produce the best overall results. Ergo, the second rule is the one we should follow, and rule pragmatism collapses into act pragmatism.
This argument will not work. After all, the rule pragmatism counsels adopting rules that will produce the best results *in the long run*. But the adoption of rules like R2 will not, in fact, produce the best results in the long run. As we saw above, rules like R2 lead inexorably into free rider problems, prisoner’s dilemmas, and so forth. And so applying rules like R2 might maximize short-term benefit, but in the long run, rules like R1 will produce better consequences.

Consider the point this way: I have argued above that the basic unit of fundamental rationality is often the strategy, not the action. But that doesn’t mean that any strategy will do; the strategy must be well-chosen. And so, for example, if my dieting strategy involves having a scoop of ice cream every day, it is unlikely that my diet will succeed. It will not achieve its long-term goal. Similarly, if the strategy chosen by the rule pragmatist is R2, then the strategy will fail to achieve its long-term goals. The strategy will be brought low by all of the problems that faced act pragmatism, the same problems that forced us to concede in the first place that the strategy is often the basic unit of fundamental rationality. Thus, the rational strategy is R1, not R2; it is more rational to select R1 than to select R2, and R2 can therefore not be said to be the choice the produces the best results.

**Rule Pragmatism is Not ‘Real’ Pragmatism**

Other critics of utilitarianism have charged that act utilitarianism is a “purer” form of utilitarianism, and that rule utilitarianism abandons what Sanford S. Levy\(^ {27}\) calls utilitarianism’s “teleological motivation”, the idea that one should begin with an idea of

\(^{27}\) Levy (1997), p. 95.
the good, and derive one’s moral theory entirely from this idea of the good. Philippa Foot puts the objection as follows:

What is it, let us now ask, that is so compelling about consequentialism? It is, I think, the rather simple thought that it can never be right to prefer a worse state of affairs to a better. It is this thought that haunts us and, incidentally, this thought that makes the move to rule utilitarianism an unsatisfactory answer to the problem of reconciling utilitarianism with common moral opinion. For surely it will be irrational, we feel, to obey even the most useful rule if in a particular instance we clearly see that such obedience will not have the best result.28

By recommending sub-optimal actions, it is thought that rule utilitarianism is “unmotivated by it or even inconsistent with it.”29 One can see how this objection might be extended to rule pragmatism.

The objection is basically the rule worship objection in another guise, and the response is the same. The objection claims that rule pragmatism is not true to pragmatism’s “teleological motivation” because it often requires sub-optimal actions. But again, this objection insists on examining rationality on the level of actions, not strategies. If we can see that the rule pragmatist strategy creates better results than the act pragmatist one, we can see that it is in fact rule pragmatism which is truer to the teleological motivation.

The Rule Specification Problem

Another standard objection against rule utilitarianism is that it often counsels foolish action when others won’t cooperate. Peter Railton, for example, writes:

---

29 Levy (1997), p. 98. Levy argues that it is in fact act utilitarianism that is inconsistent with the teleological motivation. As we saw earlier, it is act utilitarianism that fails to maximize long-term utility, by allowing free riders, by making cooperation difficult, and so forth.
I suspect that rule-consequentialism is untenable...for it would recommend acts that...accord with the best set of rules even when these rules are not in fact generally accepted, and when as a result these acts would have devastatingly bad consequences. “Let the rules with the greatest acceptance utility be followed, though the heavens fall!” is no more plausible than “Fiat justitia, ruat coelum!”—and a good bit less ringing.\(^\text{30}\)

Let me illustrate this point with an example. Suppose a car has run out of gas, and needs to be pushed to the filling station. If three people are required to push the car, and three people are present, then rule utilitarianism dictates that the utility maximizing rule in this situation would be, “Available and capable persons should help push the car to the gas station”, or something similar. But suppose that two of the three people refuse to help push. Nevertheless (the objection goes) the third person should push on the car, because this action is dictated by the rule that, if followed, would maximize utility in this situation. Thus, even though one person cannot by himself move the car, and it is utterly irrational for the third person to stand behind the car and push on it, rule utilitarianism nevertheless dictates that the lone person should push the car. Rule pragmatism might face a similar objection: if we are to follow the rules that would, if generally followed, best serve our interests, then I should follow this rule even if no-one else does, thereby wasting my effort.

This objection only has force because it assumes that the rules in a rule pragmatist system will be maximally specific. Thus, the rule pragmatist system will (goes the assumption) consist of rules such as the following:

R3: If a car has run out of gas, and a filling station is near enough so that the car can be pushed to the filling station, and there are enough able-bodied persons available to push the car to the station, then you should push the car to the station.

But once the rule is stated, we can immediately see the absurdity of assuming that a rule pragmatist system will consist of rules that are this specific. As I noted above, a moral system must be livable. And a system that consisted of rules as specific as R3 would be far too complex to learn. A more likely rule would look something like the following:

R4: If someone is in need of assistance, and you can render assistance without undue cost to yourself, then you ought to do so.

This rule doesn’t specify all the various circumstances in which one might need assistance (If a car is out of gas; if a person has a snakebite; if a person is unable to pass a test without tutoring assistance; etc.); that would result in an unlearnable profusion of rules. Nor does R4 specify how one might render help (Push the car; etc.); that too would introduce unnecessary complexity. Rather, R4 assumes that an intelligent moral agent can recognize when someone needs assistance, and crucially, that this intelligent moral person can tailor a response appropriate to the circumstances. For example, let us again consider the car that is out of gas. An intelligent moral agent, following R4, will decide that the owner of the car needs help. After trying (and failing) to enlist the help of the two bystanders in pushing the car to the filling station, the moral agent would then cast about for other ways of rendering assistance (say, walking to the filling station, borrowing a gas can, and buying enough gas to allow the car to drive to the filling station). I don’t think anyone seriously believes that a system of morality would consist of rules as specific as R3. R4 is a much more likely candidate.

How specific should the rules be? I think that there will be a trade-off between (a) the applicability and learnability of the rules and (b) the interest-promoting nature of the rules. A more specific set of rules will better foresee exceptions, be better tailored to
various, slightly different situations, and so on. But again, an overly-specific set of rules will not be learnable. And so I think a rule consequentialist will counsel us to look for an equilibrium point that represents a compromise between (a) and (b). Of course, it would be nearly impossible to sit down and construct a set of rules that satisfied this compromise. However, I think we should recognize rule formulation as an ongoing process of rule modification, keeping both (a) and (b) in mind. As such, when I speak of an equilibrium point between (a) and (b), I am not trying to give a set of truth conditions for pragmatist rules, or some ideal observer theory of the correctness of rules. Rather, I am suggesting a heuristic device that would help in the ongoing project of formulating and modifying our moral code.

In general, the response to the question, “How specific should a rule be?” is: “Whatever level of specificity best serves our interests.” If a rule is overly specific, people will have trouble learning it, will misapply it, and inefficiency will be introduced into the moral practice. But if the rule is not specific enough, then it will not display an appropriate sensitivity to various circumstances, and will hence not serve our interests well overall. For example, a rule that is more general and lumps together circumstances of sort A, B, C, and D, might serve our interests less well than a more specific rule which lumps together A, B, and C, but treats D as a separate circumstance, governed by a different rule. Again, though, questions of appropriate specificity must be settled ultimately by appeal to our interests, and to a consideration of which level of specificity best-serves our interests. At any rate, as I will argue in chapter 4, the complexity of morality precludes sitting down in our armchairs and devising a set of adequate moral rules. Moral rules must be tested by living them, and adopting, testing, revising, and
discarding moral rules is an ongoing process. In this process, there will of course be a trade-off between specificity and learnability, but this tradeoff must be negotiated over time. Thus, we needn’t specify in advance a particular balance between these two considerations; such a balance will be worked out in practice, as it must.

Having said all of this, I will concede that on occasion rule pragmatism might counsel fruitless actions. If there is an action (such as recycling) that requires a certain threshold of participation before producing a net benefit, a person might be obligated by rule pragmatism to participate even though the total number of people participating is below this threshold. I am not sure, however, that this is a fatal objection to rule pragmatism. In fact, it seems intuitively plausible that even if not enough people recycle, I should continue to recycle, in the hopes that participation by a sub-optimal number of people might eventually grow into participation by enough people. If all participants quit recycling, then there is no hope for the future of recycling; but if the current participants stick to it, then the numbers might eventually grow large enough so that the benefits of recycling are realized.

Of course, rule pragmatism would have to eventually allow non-participation, if a cooperative endeavor is obviously going to fail. But the point at which non-participation is allowed will vary according to the circumstances. Indeed, even in the face of massive non-participation, following a moral strategy can still produce desirable results: one family hiding Jews from Nazis can (obviously) save several lives, even if few (or no) other families are willing to take similar steps. In other cases, however, (such as the case of recycling) thousands must participate before any benefit is produced. So the point of allowable non-participation will vary; but non-participation will eventually be allowed.
This seems, however, like a less controversial point, and doesn’t generate obvious difficulties for rule pragmatism (other than the familiar difficulty that it will be a matter of judgment as to when the point of allowable non-participation has been reached). Thus, we can conclude that this objection seems insufficient to defeat rule-based teleological morality.

However, this objection calls to mind a similar one. What if a community is engaged in a cooperative strategy which, although it promotes our interests, is suboptimal? Should an individual who is aware of the optimal strategy (and who knows the strategy is optimal) follow the community, or follow the optimal strategy, even if others don’t?

The answer would depend on a number of factors, such as how sub-optimal the current strategy is. But even if a strategy is sub-optimal, if it is an interest-promoting strategy, it might often be rational to follow it. Let us consider a couple of examples. First, let us consider the practice of writing letters of recommendation for job candidates in academia. The current strategy is that recommenders write honest but inflated recommendations. A more rational strategy would be for recommenders to write honest and non-inflated letters. This is more rational because it would allow hiring departments to better assess the abilities of job candidates. However, it is not obviously rational for an individual to follow this latter strategy, even if would be optimal as a cooperative strategy, as doing so would unduly harm the job prospects of the candidates one was recommending.

Consider a second example. A Marxist once remarked to me that it was unfortunate that workers had to alienate their labor by working in factories, on assembly
lines, but given the capitalist organization of US society, Marxists such as himself had a moral obligation to work to ensure that workers were able to obtain such jobs and to find such jobs for workers who were unemployed. This example can be understood as follows: the Marxist believed that a Marxist economic arrangement would best-serve our interests. However, the Marxist lives in a capitalist society—one which (by his lights) follows a sub-optimal economic strategy. Now, this Marxist thinks that in an ideal society, workers would not be alienated from their labor, and would not be forced to work, say, on assembly lines. However, given that society is organized around capitalist principles, not Marxist ones, what should a Marxist try to do for workers? He ought to help them get jobs. Even though by doing so, he is encouraging participation in a cooperative strategy he does not endorse, nevertheless it might be the most rational course for him to pursue. He might agitate for a re-ordering of his country’s economic structure, but given that it is ordered in a particular (non-optimal, he believes) way, there are certain respects in which it is rational for him to participate in this non-optimal strategy.

Thus, if one’s society is following a non-optimal strategy, it might nevertheless be rational for one to go along with this strategy, rather than following the optimal one. Can we give general rules for when one should follow the non-optimal strategy? Perhaps not, but we can at least outline some relevant considerations: How sub-optimal is the strategy followed by society? Would an individual generate great pragmatic benefit by violating this strategy? What would be the repercussions (in terms of sanctions, etc.) to the individual for violating the community’s strategy? The question is not entirely dissimilar to the question of when one ought to violate a moral rule, such as the rule against lying.
A variety of considerations are relevant: How serious is the lie? Do you have a relationship of particular trust with the person you are deceiving? What is the pragmatic benefit of lying, and what is the benefit of telling the truth? By asking these questions, we can uncover when it is permissible to lie: perhaps in telling Aunt Jane that her bundt cake is tasty; definitely in telling the Nazis there are no Jews in this house; not necessarily in telling your father that you didn’t chop down the cherry tree. It is not obvious that deciding whether to violate a moral rule is itself a rule-governed activity; it is one in which practical wisdom plays a role. Similarly, in deciding when to violate and when to adhere to a non-optimal strategy, one must use practical wisdom to weigh a variety of different considerations.

The discussion of when one is justified in violating a moral rule (i.e., deviating from an optimal cooperative strategy) raises a potential objection against the current account. Let us turn to this possible problem.

Rules and Exceptions

One form the rule-worship objection might take is the idea that rules should admit of exceptions. After all, even the strictest rule utilitarian could hardly say that one ought to follow a rule with the highest acceptance utility even if doing so in this particular case would cause the destruction of all life on earth. If one rigidly follows a set of rules without regard to circumstances, and does not allow exceptions to the rule, then isn’t one engaging in a form of rule worship? But how can a rule have exceptions and still be a rule? This question arises out of the (mistaken) notion that there are precisely two sorts of rules: generalizations and exceptionless principles.
Concluding Remarks

A traditional question in ethics has been, “Why be moral?” Generally, this question has been understood as asking, “Why should I be moral?” The answer to this question is, “Because you are one of us, and it is rational for us to be moral.” Similarly, if the footballer asks why he should go right, the only possible answer is, “Because you are part of the team, and it is rational for the team to go right.”

Many puzzles in morality have been created by the insistence that rationality is act-based and individualistic. But I have argued here that rationality is in many cases strategy-based and team-oriented. Morality is precisely one of these cases: it is a case where the desired benefit can only be achieved by strategy-based team rationality. Ergo, this is the model of rationality that it is rational to apply in the moral realm; to apply another model (namely an act-based individual morality) would fail to achieve the desired result, and is therefore irrational—a failure to apply the correct rational model to a problem of means-end reasoning (namely, the problem of how best to advance our interests).

An assumption sometimes made by philosophers is that if our self-interest conflicts with our moral obligations, then it is rational to pursue our self-interest to the detriment of morality. I hope that this chapter has undermined this assumption. First, this assumption rests on the dogma of individual rationality. To assume that in such a conflict, self-interest must win out is to ignore the force of cooperative rationality, and the rational requirements it makes on us. Since morality is a rational system aimed at
promoting *all* our interests (self-regarding and other-regarding), it will often be *rational* to satisfy the requirements of morality, even if this means sacrificing some measure of self-interest. And in any case, since we have other-regarding (as well as self-regarding) interests, a comprehensive account of rationality based in interest-satisfaction will have to acknowledge that it might often be rational to satisfy our other-regarding interests to the detriment of our self-regarding interests.