IS IT RATIONAL TO CARRY OUT STRATEGIC INTENTIONS?*

MICHAEL H. ROBINS

Normally when we form an intention to perform some act at a later time our practical reasoning begins with a consideration of the action and terminates with the intention. And even though the act of intending occurs earlier than the act intended, we think that the reasons for so intending are the same as for acting. Call this the coherence thesis. The coherence thesis, however, founders in a variety of cases in which the intentions produce "autonomous effects," i.e., effects independent of and at odds with the effects of the action. Such autonomous effects often generate reasons for intending to act that are not reasons to act accordingly.

Three kinds of examples of such autonomous effects are discussed in the literature. The first, structurally the simplest, is the toxin puzzle. According to it, some rich but perverse person offers on Monday to deposit a million dollars in your bank account on Tuesday if you will simply intend on Tuesday to drink a certain toxin on Wednesday. The toxin is not life-threatening, but will make you very sick for a day or so, after which it will leave no harmful or unpleasant effects. But the million dollars that you could receive Tuesday would be yours to keep whether you actually drink the toxin or not. What reason, then, is there for you to drink the toxin? None, it would appear, but every reason to intend to drink it. And this is enough to challenge the coherence thesis. But it is important to note here and in the other cases that the thesis is challenged only if the intention to drink the toxin is sincere, not merely feigned. This does not assume, of course, that no one can ever get away with feigned intentions, but that doing so does not test the coherence or unity between reasons for
intending and reasons for acting. Apart from this, we shall return later—albeit briefly—to Gauthier's treatment of feigned intentions.

I have said that the toxin puzzle is structurally the simplest because the intention in question is categorical. But while simple, the example sounds so contrived that it appears to raise only a pseudo puzzle—or at least that is how it first struck me. The other two cases offer realism but sacrifice simplicity, for they involve conditional intentions in strategic, prisoner's dilemma contexts. One well-known example, inspired by the cold war, is the paradox of nuclear deterrence. Accordingly, each superpower has a reason to form the conditional intention to destroy another by a nuclear holocaust if one of the others launches a nuclear strike against it first. That intention is enough to deter each from a first strike—which creates a compelling reason to form and communicate such an intention. But the puzzle once again requires the assumption—this time, especially reasonable—that the deterrent intention is not credible unless each power really does intend to retaliate in the event the deterrent fails. But since this is a non-iterated prisoner's dilemma, there seems to be no reason actually to carry out the threat to retaliate once the other side does launch a first strike. This is because the reason for announcing the conditional intention to retaliate is to deter the enemy side from making the retaliation necessary. If the enemy does strike first, the deterrence objective obviously failed. What, then, could be the point of actually retaliating?

The last example looks structurally the same in that it also arises from a prisoner's dilemma but differs because, unlike the first two examples, our intuitions—at least our moral intuitions and perhaps intuitions about rationality—favor not only forming the intention, but also carrying it out. This is Gauthier's example of compliance with the intention of constrained maximization versus the policy of straightforward maximization (in Gauthier, 1986, pp. 157-188).

Again, consider a single-play prisoner's dilemma in which each person benefits by cooperating with the other. Both parties benefit by adopting the joint strategy of intending to cooperate (or as Gauthier says, 1986, pp. 157-188) having the disposition to cooperate if the other(s) does (do), but each rationally prefers to defect while the other complies with the intention to cooperate. The rational preference to