

Buridan and the Circumstances of Justice (on the Implications of the Rational Unsolvability of Certain Co-ordination Problems)¹

Duncan MacIntosh,
Dalhousie University

Published as: Duncan MacIntosh, "Buridan and the Circumstances of Justice (On the Implications of the Rational Unsolvability of Certain Co-ordination Problems)", Pacific Philosophical Quarterly, Vol. 73, No. 2 (1992), pp. 150-173.

Introduction

Many problems of rationality and morality involve agents in a partial conflict of interest: one agent can do well only if another does poorly. The paradigm is the Prisoner's Dilemma (PD). Some commentators think it rational for PD agents to compromise, agreeing to voluntarily so co-ordinate their choices that each will do fairly well, better than if each straightforwardly pursued his individual interest. Thus David Gauthier thinks that PD agents should conditionally dispose themselves to co-operate and then co-operate with the like-disposed.

But as Howard Sobel points out², in some PDs it seems impossible, directly anyway, to voluntarily comply with a compromise strategy because some of these strategies do not specify each agent's behavioral obligations. The strategy may require each agent to do something 50% of the time, for instance, but not say what each is to do this time. Sobel suggests, then, that voluntarily pursuing a compromise strategy would not always be an available alternative to say, Hobbesian Leviathan solutions, where the agents ally with a state which would then dictate actions to each agent and enforce the edict.³

I here defend Gauthier's proposal from this worry, showing how it can be rational for the agents to use randomizers to make their behavioral obligations determinate.⁴ But I then consider a deeper problem for both Gauthian and Hobbesian solutions. Both in effect reduce partial conflicts (PDs) to non-conflictual co-ordination problems (CPs). It is widely thought that rational agents can solve CPs simply by choosing by mutual fiat one among several optimizing co-ordinated patterns of choice. The only exceptions are supposed to be when the agents can't communicate, or lack certain common knowledge. Then, their only hope is that a "salient" will cause a non-rational solution. But I prove that some CPs are rationally unsolvable even absent circumstantial or informational impediments to agreement on a solution. Here, the problem is not circumstances, but rationality: As normally understood it cannot guarantee co-operative solutions to CPs or PDs, and many precepts of contractarian moral rationalism must therefore be false. I suggest some re-conceptions of rationality to alleviate some of these difficulties, but even with rationality so amended, whether rational co-operation and morality is possible is often a matter of sheer luck, not rational choice; and even such rational solutions as are possible depend on non-rational accidents.

This first part of this paper expounds Gauthier's theories; the second, a modification Sobel thinks necessary to keep them general. The third discusses some of Sobel's

supposed counter-examples to the generality of Gauthier's claims about co-operation. The last considers possibly irresolvable problems for the implementation of both Gauthian and Hobbesian systems. The larger issue is just what sorts of joint arrangements for mutually advantageous behavior rational agents can make simply by rational voluntary agreement. Resolving this question requires a new conception of the rationality of both individuals and groups. And it requires us to understand differently the rational foundations for morality as conceived by contractarian moral rationalists, who think morality is derivable from rationality (as explicated in the formal sciences of decision and game theory).

I Gauthier⁵

Gauthier aimed to resolve an apparent conflict between rationality and morality. To be rational is to maximize one's individual expected utility, to be moral, to refrain from maximizing. So how can it be rational to be moral?

Gauthier took this apparent conflict to be modeled in the one-shot PD: Each agent prefers in descending order, unilateral confession, mutual non-confession, mutual confession, unilateral non-confession. If one agent confesses, depending on what the other agent does, the first agent's best outcome is his first best, his worst, his third; if the first agent doesn't confess, his best outcome is his second, his worst, his fourth. If his actions are causally independent of the other's, no matter what the other does, the first does better (maximizes by) confessing; so if he is rational, he will confess. Ditto for the other agent. But if both confess, each achieves only his third best outcome; if only each had refrained, each would have obtained his second best.

Gauthier thinks this represents the structure of circumstances of justice, of choice problems creating moral issues and admitting moral resolutions. Moral issues are partial conflicts of interests, moral resolutions, socially optimal, mutually advantageous solutions to those conflicts. Here the agents are conflicted because neither can get his best outcome without the other getting his worst. The non-moral solution is for each to seek his maximal individual advantage, the moral one, for both to seek their joint optimal mutual advantage. This is for both to adhere to a moral rule. In a standard PD that would be: Do not confess. The defining property of moral rules: Each individual is better off if all comply with them than if all deviate, but better off still if he deviates whether or not others comply. If complying is "co-operating," deviating, "defecting," the original problem of how it can be rational to be moral is: How can it be rational to co-operate when it is individually more advantageous to defect?

Gauthier observed that while each agent has a powerful incentive to defect (confess), each, to avoid the disaster of mutual defection, has a more powerful incentive to arrange that all agents co-operate (obey the rule of non-confession). Each would prefer such arrangements even should they secure his own co-operation. But because of the ever-present temptation each has to defect, Hobbes thought the only efficacious such arrangement would be a state, an external force penalizing defectors. Each agent would then refrain from defecting to avoid the penalties, which make it maximizing and so straightforwardly rational to co-operate.

But Gauthier thinks the conflict can be resolved without the costs and coercions of the state. For he thinks agents can rationally adopt dispositions constraining them from always performing individually maximizing actions, and that an action is rational if it expresses a disposition it was rational to adopt. When facing a PD, it is most advantageous for the agent to have whatever disposition will most likely induce others to co-operate while yet allowing him to defect as much as possible. Given perfect information, this is the one which disposes one to co-operation just where, did it not, one would probably be defected against, but since it does, provokes another to co-operate, otherwise allowing one to defect. It is thus the maximizing and so rational disposition to adopt. In choosing from this disposition one will co-operate whenever one meets a similar agent, otherwise defect. Such agents are conditional co-operators or "Constrained Maximizers" (CMers), conditional because their dispositions induce them to co-operate only with the like disposed; constrained because their dispositions sometimes prevent them from performing individually maximizing actions (from "straightforwardly maximizing"--"SMing"--and so from always defecting), and instead sometimes make them perform jointly optimizing actions.

To adopt such a disposition is to adopt a strategy for making further choices, a "joint strategy," i.e., a way of choosing where one performs a certain action expecting it to be part of a certain outcome whose other part is to be a certain action of another agent. (To straightforwardly maximize--to "SM"--is to follow an "individual strategy", one where one performs a certain action regardless of what action the other agent is expected to perform.) To act co-operatively is to act as required by a joint co-operative strategy. In general, a CMer is disposed to co-operate (to base his actions on some joint strategy or practice) and will co-operate, (a) where everyone's co-operating (i.e., everyone's complying with this strategy) yields him no less utility than everyone's defecting, (b) where his share of the benefits of co-operation are at least fair to him by minimax standards⁶, and (c) where, because others are similarly disposed, his expected utility from his co-operating is higher than from everyone defecting (i.e., using individual strategies).⁷ Absent (a), there is no point to adopting a CM disposition. Absent (b), one could do better by holding out until others concede to one one's fair share. Absent (c), to co-operate would be to have been suckered, for there was no guarantee others would reciprocate.⁸ In general then, Gauthier thinks that in the circumstances of justice, each agent should "internalize a principle of action, that, followed by everyone, leads to an outcome that is both optimal and fair, in receiving the voluntary agreement of all," and that each should co-operate from that disposition with others like disposed.⁹

II Some Possible Worries

Sobel thinks it false that a rational agent will co-operate (i.e., do his part in an optimality-producing joint strategy of choice) whenever (a)-(c) are satisfied.¹⁰ For it may be that while there is a rule (or strategy) everyone's complying with which would optimize were there a clear way to comply with it, the rule is not, by itself, specific about each agent's behavioral obligations. E.g., imagine I always prefer going to the movies with you first, going alone second; you always prefer going to the library with me first, alone second.¹¹

We can't both always get our first choice. But we can do better than always going out alone by splitting the difference, each of us netting a utility intermediate between us always getting our first or second best outcomes. Each of us would be crazy to yield to the other's first choice or to demand that he do the same, more than half the time. But it is maximizing for us each to conditionally commit to the optimizing joint strategy of going to the movies together half the evenings, the library the other half. Fine. We each so commit. But where should we go tonight? The optimizing rule does not say. Since it prescribes no action to either of us, we can not co-operate. We need a randomizing process. Perhaps we can flip a coin, agreeing to abide by the result.¹² Or perhaps we can go to an official coin-flipper.¹³ The lesson Sobel draws¹⁴: a rational CMer should co-operate only given (a)-(c) and, (d) where either the joint strategy itself imposes determinate obligations on each, or does so when "partially implemented" (i.e., when it is decided for each agent how he should act, in the case at hand) by a randomizing process actually performed. Otherwise, each agent should individually SM (i.e., defect), since there is nothing he can do to act on the joint strategy.

We can distinguish¹⁵ between individual pure strategies (e.g., where an agent confesses no matter what), individual mixed strategies (e.g., where an agent confesses 50% of the time), joint pure strategies (e.g., where the agents co-ordinate, each always non-confessing), and joint mixed strategies (e.g., where the agents co-ordinate, simultaneously non-confessing 50% of the time). The trouble is that mixed strategies may not dictate actions except when "partially implemented," i.e., when how to choose has been referred to a lottery actually performed.

We will consider now two possible counter-examples to Gauthier's theses. In the first, absent an authoritative randomizer there is no dictated action and so even those committed to a joint strategy will defect. In the second, agents are committed to a joint strategy, but one the implementation of which for each agent (again, absent a lottery) depends on what the other will probably do. But then neither can individually decide what to do, so neither will co-operate.

III The Cases

Case 1

Sobel offers the following "stretched-PD" as a prima face counter-example to Gauthier's claim that rational CMers will co-operate in all PDs.¹⁶ Here, each agent prefers unilateral confession (in which he gets 100 utiles), mutual non-confession (3 utiles), mutual confession (2), unilateral non-confession (1). The "stretch" is in the wild pay-off difference between unilateral confession and all other outcomes. This makes the expected utility of taking a chance on achieving unilateral confession higher than for arranging joint non-confession; it makes it rational for each agent to conditionally commit to a joint mixed rather than pure strategy. For each has a minimax claim to 50.5 expected utilities (half of the total utilities produced by failing at unilateral confession and achieving only unilateral non-confession, plus half of those produced by succeeding at unilateral confession). The pure joint strategy of mutual non-confession only gets each agent 3

utiles, far from their minimax. Thus, by (b), neither will commit to it. Each can only get his minimax by both agents choosing by a joint mixed strategy which uses a single randomizer dictating each agent's choice, i.e., a joint randomizer dictating confession to each agent 50% of the time (or confession on a given occasion with a 50% chance). But suppose there is no randomizer. Without one, committing to that co-operative strategy underdetermines the co-operative action each should perform in nature. Thus each CMer should confess--by (d)--whatever the other may do. Thus even CMers will not always "co-operate" in nature, nor always produce optimal and fair outcomes, contra-Gauthier. Thus, he cannot, as he would like to, give alternative solutions to all of Hobbes'. In nature, and absent an authoritative randomizer, there may be no determinate co-operative choice for a joint strategy; so even CMers should sometimes defect.

Reply

It seems initially that we can make short work of these worries. First, while to co-operate we must sometimes use a lottery, the incentive to co-operate also rationalizes us in using one to implement co-operation. Thus even if there is now no authoritative lottery, it would be rational, for all we have seen so far, for us simply to agree on a lottery method, to run it, and to abide by its result.

Second, this is not us jointly allying with a Hobbesian Leviathan, since, by hypothesis, the lottery is costless and we do not need anyone to enforce compliance with it. For were it rational to co-operate simpliciter, other things equal that would still be rational where it consists in acting per the lottery.

Third, if by "nature" we mean "no co-ordination," then of course we cannot there co-ordinate. But Gauthier never thought we would. He is trying to rationalize us moving out of a state of nature in which we individually SM, into a situation where we co-ordinate, optimizing conjointly under voluntary mutual constraints.

Fourth, if we cannot find or agree on a joint lottery process, the rational alternative is not SMing and mutual defection. For we might each have our own lottery devices, and even if we can't agree to jointly abide by either, we might still be able to agree to individually abide by our personal lotteries.¹⁷ We could each resolve that instead of always confessing, we will each flip a coin to decide whether to confess provided the other is also conditionally disposed to choose using his coin. We will then each confess 50% of the time, achieving unilateral confession (and each other outcome) 25% of the time. This gives us each an expected utility for following the joint strategy of us each following the individual strategy of choosing by our individual lotteries of 26.5 (the sum of 25% of the utilities of each possible outcome). This is still a joint strategy, following it is a kind of co-operation, and resorting to it is not resorting to a Leviathan.

Fifth, if neither joint nor individual lotteries are available it is still not rational for us to SM. For we would each still do better conditionally committing to acting from the disposition to comply with the strategy of jointly non-confessing with a like-committed agent. And that is still a joint strategy, acting on it still a kind of co-operation.

Sixth, while it might be objected that Gauthier cannot accept any of these alternatives because none have an expected utility near the original minimax of 50.5, I think he could

accept all of them in their circumstances. For what the minimax point is surely depends on what options there are. If there is no way of achieving a minimax of 50.5 in the circumstances, it is not there the minimax. Rather, the minimax is whatever optimum circumstances permit. E.g., were there a machine which would furnish us each with all that we desire if we each gave it all that we have, our minimax would be all that we want. But absent such a machine it is absurd to say that in our present circumstances the minimax is "all that we want." So if circumstances permit a joint lottery the minimax is 50.5; if only individual lotteries, 26.5; if neither, 3. If no co-operative strategy is available, the question is moot.¹⁸

Lessons: First, use of lotteries is, where possible, rational if needed for co-operation. Second, it is not necessarily recourse to a state. Third, many joint strategies, pure and mixed, count as co-operation. We have not shown a PD has no co-operative solution until we have shown none work. Fourth, what counts as the optimal and fair solution to a PD depends on what the circumstances afford to resolve it.

Case 2

In Sobel's next case¹⁹, there is no mutual advantage to the agents so co-ordinating as that one agent does what the other does not. There is mutual advantage to each doing the same, but differential advantage to each depending on which of two possible same-type actions are performed. An example would be a PD where one agent's preferences would have him get 3 utiles if both non-confess, 2 if both confess, 0 from everything else; the other agent, 3 from if both confess, 2 if both non-confess, otherwise 0: one agent gets more utility if both non-confess, the other, if both confess. Thus they differ on which joint solution they favor--both confessing or both non-confessing--though each favors either over individual strategies.

Reply

This case presents several difficulties. First, while CMers would commit to either each of them confessing or each not, that commitment does not say which action each should perform; should each confess, or each not? And neither can act on the joint strategy until this is settled. Without a lottery it seems then that they cannot implement the joint strategy, so neither agent can act co-operatively. (Couldn't either agent provoke the other to do his part in the strategy most favored by the first agent by the first agent doing his part in his favorite strategy first? For then the second agent always does better going along with the first agent's preemptive move than not. Better that both do the same thing, even if one agent has usurped the prerogative of deciding what it will be. Unfortunately, if both think this way, as they must if both are rational and if it is the rational way to think, each will race to make his favorite choice, attempting to force it on the other, the result, a failure of co-ordination. Since both agents are presumed to be well-informed and rational, both would see this and its consequences, and both will pause and attempt to arrange non-preemptive co-ordination.) Even were both agents committed to a joint strategy, to choosing like each other, and even if they sought to resolve the dilemma

without a lottery, by fiat, neither agent could rationally decide how to choose unless he knew how the other was likely to choose. Since nothing makes it more rational to choose one way over the other for either agent, neither can ever make a rational choice. But a lottery would clearly resolve this. And since a lottery would be instrumental to a strategy to which it is maximizing to commit, everything equal it would be maximizing and so rational to commit to compliance with a lottery.

But there is a differential advantage to the agents in the outcome of the lottery. Thus if it says both must confess, the second agent gets one more utility than if both don't confess and vice versa. What is the minimax? How can a joint strategy attain it? Ideally the agents would commit to splitting the advantage, their minimax an even split. Were that impossible, there is still no problem, for the initial randomization of behaviors distributes the advantage, equalizing expected utilities, and while one agent must gain more actual utility, the other cannot object. For there is no other available more fair procedure, each had equal chance at the advantage, and without this process the "looser" does even worse.

But what if no joint randomizer is available? Then there can't be co-ordination. (Well, not quite. Suppose that, though the agents had no agreed joint randomizer, they each had individual randomizers. It would advantage each agent to conditionally commit to the joint strategy of each agent choosing whether to confess or non-confess by each using his individual randomizer, for then they at least have a 50% chance of attaining one of the equilibrium outcomes, and each agent has a 25% chance of attaining his most preferred outcome. While if each simply followed the individual strategy of choosing his half of his most preferred outcome, each would have no chance of attaining either equilibrium outcome. But of course, it could well happen that neither equilibrium results in the outcome produced by these individual randomizations. In any case, let us stipulate that the agents also lack individual randomizers.) But then neither would Gauthier think there ought to be co-operation (i.e., successful co-ordination in producing one of the equilibrium outcomes). For here (a) and (c) are not satisfied: No (available) joint strategies would maximize if agents chose on their basis over SMing; no action is advantageous over the one the SMer would perform were all others SMers. So Gauthier need not solve this case any more than he must solve PDs where agents are not given pre-interactive information on each others' dispositions, nor a pre-interactive chance to amend them. It is not a circumstance of justice; unsurprising, for a crucial requisite to its being one--a joint randomizer or individual randomizers--is stipulated unavailable. It is not the game's pay-off structure which prevents it from being a circumstance of justice, but that absent a lottery (or lotteries), even with the best will in the world the agents cannot attain an equilibrium outcome, cannot choose expecting a minimax higher than for individual strategies.

But is Hobbesianism superior here? One might think so. For Hobbes might tell us to ally with a Leviathan who would then impose a solution. The Leviathan might use a lottery available only to Him to decide whether to impose confession or non-confession, and if we knew this, it would individually maximize for us each to conditionally commit to gambling with the Leviathan. Or He might simply have an arbitrary, secret policy of enforcing, say, non-confession. It would, again, advantage each of us to take a gamble and ally ourselves with Him over us SMing no matter what his policy, even though, one

of us will be marginally disadvantaged by it, whichever it is. (His policy must be secret; otherwise, if it is a condition of His existing as a Leviathan that we both ally with Him, the agent who knew himself to be marginally disadvantaged by the Leviathan's policy would have a stake in holding out, breaking the deal. Nothing would rationally compel his acquiescence in the Leviathan's solution if his preferred solution could still be attained by the other agent's agreeing to go along with it. Rather, the old problem returns: should the agents go with joint confession or joint non-confession? But now the problem takes this form: should they go with the Leviathan's policy or make their own pact?)²⁰

But this is not really a fair alternative to Gauthian solutions. For one thing, just as the Leviathan may have an arbitrary policy, so might one of us (one either hard-wired or involving an additional preference). Oddly, it would be to our mutual advantage if one of us was known to be incapable of acting in any way other than the way that would involve compliance with the strategy favoring him. Paradoxically, sometimes incorrigible obstinacy enables agreement, co-ordination and compromise. For it would advantage us to agree to abide by the arbitrary policy one of us has so as to avoid the overhead costs and coercions of Leviathans. Our private commitment need not consist in accepting a state; for there need be no exogenous penalty for defection, nor any additional cost to the solution.²¹ Of course, had we each a policy bias we could make no progress: who should (could?) concede? But this brings me to the other thing.

To present Leviathan solutions with a problem truly parallel to the one we posed for Gauthier, we must imagine there being two (or two possible) Leviathans, one, known to enforce confession, the other, non-confession. It is mutually advantageous for us to each commit to the same Leviathan, but which? It is not uniquely rational for us to both make a Hobbesian side-bet with either; each of us would favor joint alliance with a different Leviathan. Indeed, before we can rationally co-ordinate on a choice, we still need a joint randomizer to choose. Hopefully a third Leviathan would randomize to decide which policy to enforce, in which case our individual expected utilities would be highest allying with him compared to SMing. (Each of us would prefer us to ally with the Leviathan whose policy is known to favor us, but neither of us can rationally expect the other to accept this over gambling with the randomizing Leviathan.)

So even if Gauthier can't give a co-operative solution here (because of the unavailability of a joint randomizer), neither could Hobbes give his political one. This proves that recourse to an authoritative lottery need not be recourse to a state. For in attempting a Hobbesian solution to some PDs in nature, we must use an agreed-to-be-authoritative joint randomizer to launch it. We cannot rationally decide to which Leviathan to commit without one, or unless one of Them has one. And what makes one randomizer authoritative is not that a Leviathan enforces the results of His randomizer, but that we rationally acquiesce in taking His to be authoritative. Thus the randomizer is a device for getting from nature to a state. It must thus be available in nature, not only as an alternative to the decision-making processes available in nature.²²

Thus the requisites for rational joint alliance to a Leviathan resemble those for rational joint commitment to a strategy for further voluntary choices. At least one of two possible requisites must be available. One is that of a joint randomizer. The Leviathan must be disposed to use one to decide His policy if we are to rationally ally with Him; the

contracting agents must use one to implement joint strategies for further choices. If one is unavailable, the requisite is that there be only one available mutually advantageous policy--the Leviathan must have (a secret) one; or, in a Gauthier solution, exactly one of the agents must have one (one which he cannot but insist on, forcing the other agent's acquiescence). This is unsurprising, for Gauthier's solution to PDs was to be like Hobbes' in two ways: First, the agents each choose it for individual advantage. Second, when the agents severally choose it, optimality is the by-product. But Gauthier's solution needs no coercive and costly enforcement. It can offer moral (and rationally attractive) solutions to more conflicts than could Hobbes', for in being cheaper it will more frequently surpass individual strategies in expected utility. But there are deeper problems here....

IV The Rational Unsolvability of Co-ordination Problems in Mixed Joint Strategies; Reflections On Luck

There is a deep issue near the first two cases, most clearly near the second: Is it always in principle possible for agents to rationally adopt a lottery whenever one is needed to implement any rational joint strategy? I have so far treated these cases as unproblematic because it seemed obvious Gauthier's principles would naturally direct agents to lottery solutions to CPs, and that only some unfounded stipulation would make that option unavailable. But if there are systematic difficulties with adopting lottery solutions, these cases become surprising counter-examples to the scope of Gauthier's claims concerning the rationality of co-operation in PDs. Moreover, if this does not plague Hobbesian solutions, Gauthier would have failed to offer a lower-cost alternative to all of them. This worry is not much developed in the literature. While it is unclear exactly how authoritative joint randomization is to be undertaken, it is often assumed that it can, perhaps, be done by the agents' mutual fiat. But I will now argue that it cannot always be done in this or any way.

Ideal PD agents would see the advantage in deciding their choices by a joint lottery; also in jointly baptizing one of the many possible lottery methods as authoritative if none are already forced on them by a state or by there being but one lottery method available in nature (in which case they would so baptize it); also that if several methods are available, it would optimize for them both to commit to one; also for someone to run it. Still, some vital questions remain unanswered²³: If several lottery methods are available in nature, how are agents to rationally and jointly choose one as authoritative, and to decide who shall run it? The rational mandate to use one does not say which one to use, nor who should run it. If these problems have no determinate solution the co-operative strategy cannot be (rationally) implemented, and, where only a joint mixed strategy beats individual SMing (as in Case 2), no co-operative solution is possible.

I see no way to solve these problems in the standard conception of rationality. Should we flip a coin to decide which randomizing process to use as authoritative (e.g., a further coin-toss, or some natural, random phenomenon), then flip again to decide who shall run it? But this initiates a vicious regress since presumably we need yet more lotteries to decide which one shall determine the authoritative one and which to use in determining who shall run the one for determining who shall run the authoritative one, etc. Hopeless.

Maybe one of us should just volunteer. Fine, but who? It doesn't matter. Then how are we to decide? Just do it. What if we both propose to just do it? One must yield. Who? It doesn't matter. Right, so who? OK, the agents should take the first available method; the agent closest to it should run it. But why? And what if two methods turn up simultaneously, or if the agents are equi-distant to the single method, etc.?²⁴

If we are lucky, only one process will be available (or the others will be costlier), and only one of us will be able to run it (or the cost of the other doing so will be prohibitive). Since it would optimize for one process to be run by one of us, and since there is only one possible (or cost-effective) process and one possible (or cost-effective) operator, it would maximize for him to operate it: For he cannot have the benefits of co-operation unless he does so, while by doing so, he can unilaterally save himself from mutual defection (this incidentally benefitting his partner). That is, it would maximize for him to unilaterally commit to the strategy of implementing the lottery process necessary to implementing the joint strategy (providing this is not more costly than his benefits in having it implemented). And it would maximize for the other agent to conditionally acknowledge this lottery as authoritative, and to conditionally commit to complying with it.

We might reflect here that while this problem is vexing, requiring the non-rational good fortune of a paucity of methods of co-ordination to facilitate implementation of the co-operative strategy, that we now face this problem betokens tremendous progress: We have reduced a PD to a mere CP. We now see the advantage in co-ordinating and no longer see ourselves as conflicted. But there is still the CP, every reason to think some CPs are rationally intractable. For while we do not care which lottery method we use (if they are costless or equally costful), nor who runs it (if that is costless or equally costful for both of us), this is just what makes it impossible for us to rationally choose here; and barring happy accidents limiting our options to one, it seems we cannot (rationally) resolve these difficulties. We will then, wherever an optimal joint pure strategy is also unavailable (e.g., in Case 2), be unable to co-operate.²⁵

IV.1 Rationality and Equal Options

What prevents a solution is that the agents' indifference between methods means they can't choose one. But doesn't this presuppose that an agent cannot rationally choose between A and B unless he prefers one to the other; that two agents cannot jointly choose between A and B unless both prefer the same? But surely in choosing between good A or good B and bad C, an agent who prefers A or B to C and who chooses A or chooses B--no matter how--chooses rationally. He is rational if he chooses A over C, or B over C, and is at worst non-rational, not irrational, in choosing A over B or B over A. So an action is rational if there is at least one true description of it on which it is prima face rational (or better, at least one partitioning of options where a choice between them is rational), as here. Further, since nearly every possible action must contain components from options between which the agent is preference-indifferent (e.g., the exact shape of the letters he uses in signing a check), beings capable of ordinary actions must have things like deterministic or randomizing choice tendencies to "fill in the details" of rationally preferred actions and enabling choices between equal options.²⁶ So it would be

absurd to call an action irrational simply because, perforce, it involves non-rational choices in its details. So "choosing" a lottery method is not irrational just for involving these.

But can someone be said to choose an option he does not prefer over all others? For then he doesn't choose it in the sense that it comes to be his by a preference that only it be his. It just somehow comes to affect his behavior or utility, which seems to be something that merely happens, not that he makes happen. But let us be liberal: an individual has chosen an option just if (1) it affects his subsequent behavior (or utility) because he preferred it over all others (i.e., it is "preference rational"), or (2) it does so partly because he preferred it or other options over some he dispreferred and partly because of some non-rational process (i.e., it is "process rational"). Thus a rational action is not just a preference-expressing or satisfaction-maximizing one. Rationality is dirtier: to maximize one must often first, or componently, be able to choose among equal options. Rational choices are constrained but not fully determined by preferences and beliefs.

But does granting any of this about rational agents make any difference to whether they can co-ordinate out of nature? Suppose it would advantage two such agents to co-ordinate their choices by any of many available further methods of joint choice, but that each is indifferent about which to use. Can they always in principle choose one? Obviously not by mere preference. Can they with their several individual further choice processes? Only on one of four conditions: (i) By co-incidence these processes are the same in kind and so would always choose the same options. (ii) While they are different, one always automatically acquiesces in the other's choice, the other always actively chooses; their processes are complementary. (iii) By a chain of choices involving individual processes and preferences (e.g., a series of "binary sorts"), their further processes could be made the same or (iv) complementary, for any case. Conditions (iii) and (iv) are just special cases of (i) and (ii) respectively, and (ii) is functionally just a special case of (i).

There are only two possible ways for (i) to be implemented: by deterministic or randomizing processes of further choice. Suppose both agents have deterministic ones. They may try to agree on which lottery method to use in choosing their actions in a PD by individually applying their processes in a series of binary choices between all available methods. Assume there are finitely many. Assume each agent arrives at a single method. If they still differ they need only agree on which of the two remaining ones to use. Now their individual processes, in individually choosing among these two, will either choose the same method or not. If the same, they can use it to solve their CP; if not, they cannot solve it. Since the former is not guaranteed, they may be unable to solve it.

Consider now agents each with randomizing processes of further choice. They individually choose as before, each arriving at a method. Now they only need agree on one of these if they still differ. Can they? If they each have a randomizing process that selects a given option in a choice between two with a 50% probability (though any non-zero probability would be helpful), then on their first attempt at agreeing, they have a 25% chance of succeeding. (Success is each of their processes voting yes for a method; thus on any try, there are 4 possibilities, yes/yes, yes/no, no/no, no/yes, the chances of yes/yes, 1/4.) By every successive attempt they have an ever higher chance of having

agreed. Were this series finitely convergent, they would be guaranteed to agree in finite time (if each try takes only finite time). Unfortunately, the series only asymptotically approaches unity. So while they are guaranteed consensus in infinite time, there is no guarantee they can solve their CP by any given time.

Since every possible choice process is one of these two, and since neither is guaranteed to solve their CP in finite time (all that they ever have), no method is guaranteed. Since they could only rationally co-ordinate if they could solve their problem, and they may not be able to, rationality cannot itself always assure co-ordination in a PD. Nor therefore can it, by itself, always rationalize co-operation. Co-ordination between the agents is possible only should they be lucky enough to have co-ordination-facilitating processes lucky enough to solve their CP in the given time. So whether they can rationally co-ordinate and co-operate depends on luck; on something beyond the agents' rational choice as defined by their individual preferences or choice processes. So even granting that an option can be rationally individually chosen where it could not be preference-chosen over equal options (and even if a joint choice is rational if it results from individually process rational choices), it still does not follow that rational agents in nature can always co-ordinate by fiat. No matter how rational they are, nor how willing to co-operate, they will not always co-operate in all possible PDs where the only obstacle is their agreeing on a co-ordinating device.²⁷

But suppose we insisted that truly rational communities could always resolve such CPs (--for wouldn't we think a community was crazy if its members just left crucial issues hanging because they couldn't overcome such "trivial" obstacles to agreement on how to act?; and isn't it the business of rational choice theory to recover from our pre-critical judgments about what is rational and what isn't, some generalized principles summarizing and explicating our pre-critical "intuitions"?)²⁸: Then we've just proved that the rationality of a community (i.e., the ability of the members of the community to coordinate in ways they believe likely to achieve agreed goals) is not the sum contribution of its members conceived as merely preference-rational, nor as also process-rational. It requires a further lucky natural suitedness to co-ordination among their processes. We can put this implication in another way. Suppose that strategic rationality concerns choices where outcomes depend on the choices and/or choice bases of other agents, parametric rationality, choices where outcomes depend only on natural conditions and the choices and/or choice bases of the individual chooser. Then we have shown that the strategic rationalities of agents in a community of choosers differ from parametric in demanding not only choice processes, but also luck in them being suited-to-co-ordination with the individuals in the community. That is, two or more agents are not strategically rational in their dealings with each other unless their several choice processes luckily suit them to co-ordination with each other.

IV.2 Leviathans

Suppose it is to our advantage to co-ordinate, but that we happen not to be endowed with suitable co-ordination-facilitating processes. Would a Leviathan help? No. Hobbes thought a Leviathan could be established in either of two ways, by "institution" or by

"acquisition."²⁹ In the former, the Leviathan comes to exist by agents in nature creating one to escape the vicissitudes of nature: it is supposedly rational for a given agent to ally with a Leviathan if all others are prepared to do so as well. The agents have as their only incentive to so ally that their joint commitment would remove them from nature and place them under the Leviathan's protection, freeing them from the war of all against all. The only penalty for not signing up is the continuing nastiness of the alternative--life in the state of nature. There are no further individual signing bonuses, nor further individual penalties for not signing. After they signed up there were rewards and threats aplenty from the Leviathan to enforce His policies, but that's different. Now, unfortunately, establishment of a Leviathan by institution involves co-ordination. And if the agents were not guaranteed of being able to co-ordinate on a lottery, then they are not on a Leviathan either. For they must agree on who He shall be, and on what policies He will have. And it was just their inability to agree on policies which required them to resort to a lottery in Case 2, their hypothesized inability to co-ordinate on which lottery to use that drove us to consider whether a Leviathan could help. But as we've seen, we may be unable to rationally agree on joint alliance to a single Leviathan because unable to agree on which policy He should enforce. While if we propose to institute a randomizing Leviathan, we have still to choose among equally attractive possible Leviathans with different equally attractive lotteries; and we could not rationally ally with either if we could not co-ordinate in general in choices between equal options.

It is surprising how long this problem has gone unacknowledged in the literature. Consider a recent book-length treatment of Hobbes's solution to such problems. Jean Hampton³⁰, recognizing that agents in nature face the problem of selecting a Leader, a Leviathan, suggests that that problem can be solved using one of the following methods:

- a) One of the agents could impose additional costs on the other should he refuse to accept the Leviathan the first prefers. But what if, as I stipulate, the agents have equal powers to impose costs, so that either or neither can make the other's refusing to concede so expensive as to induce the other to capitulate? Then either both will be willing to capitulate and they will have their problem back in deciding whose capitulation shall stand, or neither will capitulate, and again, their problem returns.
- b) One of the agents can hold out until the other gives in. But what is to prevent both from holding out? And what is to suddenly make it rational for (just) one to give in?
- c) Since each has a stake in some resolution being attained, each has an incentive to concede to the other's preferences. But since both have this incentive, both should be prepared to concede; how are they to decide whose gracious concession to accept?
- d) Since the problem becomes more urgent as a deadline for a solution approaches, one agent should concede as it approaches. But while the approach of the deadline makes things more urgent, given that the agents have equal stakes in their preferred solutions, who should rationally cave in?
- e) The game is like a game of chicken, and at the last second, both should deviate from the deadlock. But how are the agents rationally to decide who, as it were, shall veer left, and who right?; what in rationality can say who should concede and who should hold?
- f) One of the agents should yield to the other's irrational inflexibility. But what if both agents are perfectly rational?; what if neither is particularly inflexible?
- g) One of the agents might have a greater stake in achieving a resolution, and so should, rationally, be more willing

to concede. Sure, but what if they have equal stakes, as, I stipulate, is the case here? h) The more risk averse agent should cave in. But what if, as I stipulate, the agents are equally risk averse? i) The agents should use a symmetry-breaking technique, like flipping a coin. Right, but what if they each have a coin, etc., as I stipulate here? How are they rationally (or even by fiat) to decide which of several equally attractive symmetry-breaking methods to use? This just returns us to our original problem.³¹ j) The agents could vote on a solution. But when the problem holds between two agents, as here, a vote may not settle the question; voting is like them each choosing a joint lottery method by their several further choice processes, and there is no guarantee of the agents being able to reach a consensus on the joint lottery method in a finite time using these processes. Even with all of this flailing about, then, there is, I insist, no guarantee that the agents can rationally solve their co-ordination problem in attempting to institute a Leviathan. All of these proposals depend on contingencies in the provision of which lady luck might not have obliged.

But couldn't a Leviathan solve our problem for us by imposing Himself on us, forcing us to so ally, and so resolving our CP on choosing a method of choosing a solution? This involves the other way for a Leviathan to be established, vis., when one acquires the loyalty of a subject by conquest. It is "be loyal or else." Since people can be "conquered" by positive as well as negative reward, perhaps "if you're loyal I'll give you this" would also count as establishment by acquisition.

Now I have three points to make about this. First, nothing guarantees it to happen. Hampton³² assumed that where there were several candidate Leviathans, one will be the strongest, and the agents, to avoid the harms of this Leviathan and to protect themselves from other, weaker Leviathans, and from each other, would ally with it. But what if there are equally strong incentive-or-threat-offering competitor candidate Sovereigns? How are the agents to pick? Discussing this, Hampton assumes³³ that one confederacy (Leviathan) will win out in the "market" of confederacies as the strongest one, making it the one with which the agents ought rationally to ally. But what ensures this? Why might there not, again, prove to be two equals? Again, considering this scenario³⁴, Hampton claims that the agents coming to have a Leviathan by acquisition is really the agents making self-interested choices, co-ordinating by means of a salient. They want a Ruler, both want to have the same Ruler, whichever one they end up with, and both are happy to have one chosen for them by that Ruler offering the strongest incentive to alliance; that Ruler's incentive makes Him a natural "salient." (Thus she seems to think that a Leviathan by acquisition reduces to a Leviathan by institution.) Fine, but now recall our old problem. Suppose there are two candidate Rulers, offering equal incentives; who do we want to conquer us? Suppose either of these Rulers would become strongest if we both allied with him, but each is equally attractive for each of us, or each of us mildly favors a different one: how are we supposed to solve our problem? Contra Hampton³⁵, it seems that nothing will inevitably lead to one Ruler's having a monopoly on ultimate power, and so there is nothing to inevitably make one Ruler naturally "salient," the one with which it would be rational for both agents to ally. Hampton seems wrong in saying that "nothing is required for the sovereign's institution that these self-interested people are unable to perform."³⁶ In fact, there may be nothing they can do to make any candidate Leviathan the one with

which it is rational for both agents to ally.

My second point on Leviathans by acquisition: Some of the ways Leviathans acquire subjects do not happen by the agents' co-ordinating. Rather, agents individually ally with the Leviathan to escape an individual penalty or to attain an individual reward. They pursue individual, not joint strategies.

My third point: In those ways in which acquisition does happen by the agents' co-ordinating, it does not happen by them co-ordinating in a way that can count as them having solved their original choice problem. The situation may be one involving co-ordination of some form, as where the Leviathan's threats and seductions induce co-ordination ("You must both comply with my wishes; if either of you does not, you both shall be punished/neither rewarded"). Here, the agents solve a problem by rationally co-ordinating, but not their original one, only the new one of how to escape the wrath (or earn the rewards) of Leviathan. None of this is a "solution" to the original choice problem because it involves exogenous factors, ones external to both the original scenario and its pay-off structure. The Leviathan changes the problem into one in which there is a unique best equilibrium, so that the agents no longer solve their problem by a selection by mutually agreed fiat from among a plurality of good equilibria--there is no longer such a plurality to select from. The agents have not solved their problem; rather, an external force has come in and changed their situation so that they no longer have it. This can of course happen, but it is not a rational solution to a partial conflict as such, since it removes the conflict.

To summarize then, a Leviathan by acquisition is not guaranteed to present itself, and even if it does, opting for it is not a solution to the original choice problem (but to a new one). While a Leviathan by institution is no more guaranteed to work as a strategy of solution than attempting to resort to a joint lottery. Unsurprising, for the device of an instituted Leviathan was invented to solve a different problem than that of agreement on a co-ordinated pattern of behavior, vis., "how can it be rational to comply with such agreements?", not "which agreement (on how exactly to co-ordinate) is it rational to make?"³⁷

Agents unlucky enough to be insufficiently endowed as to be able to co-ordinate can only escape (not solve) their problem by the intrusion of an exogenous factor. Whether a genuinely co-operative solution is possible in a given circumstance of justice (or whether something is such a circumstance) depends in part on sheer luck (in us being co-ordination suited); and we can't overcome bad luck by reason alone. This reduces the number of circumstances admitting of a Gauthian resolution. In effect, morality (and the solvability of a co-ordination problem) depends on luck in the very operation of rationality. (This is a systematic problem. Of course we must also be lucky enough not to be blown up by a volcano in the meanwhile.) Ironically, reasoning our way out of partial conflict through rationally instituted moral or political artifices, systematically depends on non-rational accidents.

But it would be wrong to think that wherever a Gauthian resolution must fail, the Hobbesian solution of mutual rational alliance to a state can succeed. For we saw that rationality may be unable to institute a Leviathan to extract us through rational commitments from the worst circumstances of indecision. A solution could only arise

non-rationally (relative to the original circumstances of choice) from natural accidents (suiting us to co-ordination); or an escape (not a solution as such) could be afforded by the intrusion of a (single) conqueror with his own agenda. Thus that venerated artifice, the Hobbesian state, once presumed to attract our rational allegiance as our last recourse from the war of all against all, must often be a purely natural object, one in principle not rationally institutable by its prospective constituency; it must have its own agenda.

We nearly always have many methods and agents of randomization. Since our rationality affords no a priori guarantee that we can always in principle rationally jointly choose among them, how do we so frequently manage to co-ordinate and co-operate? Evolution would favor communities of agents either genetically endowed with convergent tie-breaking processes, or who were not preference-indifferent and who shared preferences on these matters (or who were, happily enough, subject to certain kinds of conquest). Evolution favors the prejudiced and the picky (and sometimes the conquered), both in individuals, and, especially, in communities. For such communities could solve CPs to their advantage, while others would be stuck with sub-optimal individual strategies. Our present co-operative practices, then, are the beneficiaries of the evolutionary tendency to produce individuals suited-to-co-ordination. But as we saw, the possibility of co-ordination, and so of co-operation, cannot be guaranteed by the individual rationality of agents. It is an unwitting grace of forces themselves of necessity non-rational.

Conclusions

We can now see that some common precepts of contractarian moral rationalism are false. That a rational choice is a unique function of an agent's preferences given his beliefs was falsified by the fact that an agent can face a choice among equal options. If he can rationally choose among them, rationality must merely be constrained, not exhausted by, the expression of preferences and beliefs; there must also be what I called process rationality, not just preference rationality.

That the rationality of collective action (i.e., of groups of individuals co-ordinatively interacting to attain agreed goals) is a unique function of the rationalities of the individuals comprising the collectivity was disproved for both individual preference and process rationality. For were individual rationality merely preference rationality, then just as individuals can face impossible choices among equal options, so can groups of individuals; a group can be indifferent among options even where it prefers at least one of them to some inferiors. And even supplementing individual preference rationality with process rationality so that agents can break preference ties in their individual choices, groups can still face inco-ordinable choices among options process equal for all individuals. The processes individuals use to break ties may not combine to allow groups of individuals to do so. If we hold that groups can always make rational such choices, we must understand the rationality of a collectivity not just as the expression of the preference rationality of its individual members, but as that plus their process rationality, plus a further constraint on the latter to suit the individuals for co-ordination. Moreover, since nothing guarantees that any set of individuals will have several choice processes

jointly sufficient to enable agreement on how to break ties in collectively designing such processes, this suitedness to co-ordination cannot itself always be something designable by the group. Rather, it must sometimes be in some way non-rational and adventitious.

These reflections on individual and group rationality suggest that to understand strategic rationality (the rationality of agents in interactions with other agents aiming at outcomes which depend on the choices of several agents), we must take groups as our unit of analysis, not individuals; individual strategic rationality must be understood as an irreducible relation (of suitedness to co-ordination) between the individual and her enviroing group, not as a trait of an individual considered separately. Our reflections also suggest that to understand the rationality of a group (i.e., the ability of its members to co-ordinate in attaining agreed goals), we must invoke evolutionary forces, ones bearing on the group and so on individuals in their relations to their group. We cannot suppose the group's rationality to consist simply in the sum of the preference and process rationality of its members taken separately.

Further, since we must understand the strategic rationality of an individual in relation to her enviroing group, our analysis may suggest that whether someone is strategically rational depends on what group she is in. An agent may have processes of choice suiting her for co-ordination in one group, but not in another. So whether she is strategically rational may be relative to what group she is in. But we might go even further: Suppose we thought that for any two people (regardless of their native group), the two of them are only strategically rational if, in possible dealings with each other, they could always co-ordinate in the kinds of situations we have considered. Then it is a condition on any agent's being strategically rational that all beings which can count as agents are luckily such that their choice processes suit them to co-ordination.³⁸

That optimal resolutions of partial conflicts are always achievable by rational mutual compromise was disproved by the fact that the stubbornness of one agent where agents must choose between otherwise equal options may be the only thing allowing them to solve their CPs. The rest must defer to he who cannot yield, his inflexibility useful in eliminating otherwise equally attractive options.

That the state can be understood at least hypothetically as a wholly designed object chosen by its subjects was disproved by the fact that there may be two equal Leviathans so that, since the group could not preference or process rationally choose between them, the only way it could have come to be ruled by one is if it imposed itself.

That a just--welfare optimizing--state, must be choosable by a community, fell with the fourth dogma: there could be an optimizing state a group could not rationally choose. And finally, that any maximally just state must be one to which all its citizens would hypothetically commit in individually rational contracts was disproved by the fact that several states could offer optimality, and individuals as rational contractors might not, therefore, rationally be able to contract for even one of them.

Notes

1. I began thinking about the issues discussed here when I was the commentator on a paper Jordan Howard Sobel presented at the 1989 meetings of the Canadian Philosophical Association. A descendent of his paper has appeared: "Constrained Maximization," Canadian Journal of Philosophy, 21 (1991), pp. 25-52. For helpful comments made variously on my commentary and on a draft of the present work read at Dalhousie University, my thanks especially to Sobel (who gave me written comments) and also to John Baker, Nathan Brett, Steven Burns, Richmond Campbell, Bob Martin, Jan Narveson, Sue Sherwin, Terry Tomkow, Jane van Arscotte, Peter Valentine, Kadri Vihvelin, Michael Webster and Sheldon Wein. Thanks also to the anonymous reader for this Journal for many helpful suggestions.
2. Sobel, "Constrained Maximization."
3. Sobel, "Constrained Maximization," pp. 25, 37, 40, 42.
4. I here develop possibilities Sobel negatively reviews in passing. See Sobel, "Constrained Maximization," pp. 28, 29-30, 48, 50.
5. Parts of the next five paragraphs are loosely adapted from my 'Preference's Progress: Rational Self-Alteration and the Rationality of Morality,' Dialogue, 30 (1991), pp. 4-5.
6. Little hangs on the details of minimax relative concession here, but briefly: If we each invest equally in an enterprise, then, ceteris paribus, we should profit as much and as near to equally as possible in the circumstances.
7. David Gauthier, Morals By Agreement (Oxford: Clarendon Press 1986), p. 167.
8. Gauthier, Morals By Agreement, p. 168.
9. David Gauthier, "Moral Artifice", Canadian Journal of Philosophy 18 (1988), p. 389.
10. Sobel, "Constrained Maximization."
11. This is a version of an example of Sobel's (Sobel, "Constrained Maximization," p. 28), in turn adapted from Gauthier.
12. David Gauthier, Morals By Agreement, p. 120.
13. Sobel, "Constrained Maximization," pp. 29-30.
14. Sobel, "Constrained Maximization," pp. 28, 30, 33.

15. As Sobel does in more careful and technical terms, in his "Constrained Maximization," p. 27.

16. Sobel, "Constrained Maximization," pp. 34-37.

17. Jan Narveson suggested the individual randomizer. I am not sure whether he meant that the agent should pursue an individual strategy of individual randomization, or that each agent should provisionally commit to a joint strategy involving each agent's agreeing to the individual use of individual randomizers. I presume the second, where each agent only randomizes provided the other conditionally disposes himself to do so too, for the first is irrational. If I randomized, while you straightforwardly maximized, that would have you always confessing but me not confessing half the time, so I would be getting suckered half the time.

18. Perhaps the outcome utility of mutual defection is then the minimax: for neither agent would rationally accept less, nor could he rationally expect more.

19. Sobel, "Constrained Maximization," pp. 38-43. I've simplified his case a bit.

20. Compare with Gregory S. Kavka, Hobbesian Moral and Political Theory (Princeton University Press: Princeton, New Jersey, 1986), pp. 182-188, especially pp. 186-187.

21. For that matter one of us might just happen to have a fair randomizer. And it would then advantage each of us for us each to duly conditionally commit to taking it for the authoritative one and to complying with its dictates, again saving ourselves the overhead costs and coercions of the Leviathan.

22. Except in the sense that from an unconstrained situation, we can rationally move to mutually agreed constraint by agreeing to refer the choice problem to an agreed joint lottery, not to be confused with an agreed coercive force.

23. I became aware of these questions through Sobel, though he does not put the problem in quite these terms.

24. The reader for this Journal offered the following illustrative example and commentary:

Two people need to go through a narrow doorway. One must go before the other, it doesn't matter which one. So who should go? They can easily communicate. One says: "After you," but simultaneously the other is saying "After you." At this point both start going through the doorway. This is clearly unsatisfactory. So they both stop. Then each says "I'll go first." One can see that in principle this could go on forever with them never getting through the doorway, though no irrationality is in evidence...even when communication is possible and an explicit agreement can

be made, rational agents will not always be able to make such an agreement.

Exactly. A more detailed proof follows in the next section.

25. Gauthier agrees that where there is more than one best equilibrium, "coordination is a matter of chance, not of reason." (David Gauthier, "Coordination," Dialogue 14 (1975), pp. 195-221.) But he thinks that "the salience of one of these equilibria...may be used to reconceive the situation as having but one best equilibrium." Successful coordination is possible if "all the persons involved in the situation apprehend the same outcome as salient." (p. 213) Perhaps, though Margaret Gilbert in her "Rationality and Salience," Philosophical Studies 57 (1989), pp. 61-77, argues that unless one of these options being salient alters the pay-off structure in utilities of the choices, however good for both agents it would be if both chose by the salient, this does not give either agent an individual reason to choose by it, even if, as a matter of contingent psychology, and independent of pure rationality, agents might both tend to choose by it (pp. 71-73). In any case, if we stipulate, as I here do, that there is no salient for one of these options, or if several are equally salient, Gauthier will acknowledge that "coordination is a matter of chance, not of reason." And he admits there is no guarantee of a unique salient ("Coordination," p. 211). He does not seem to have fully traced out the limitations this may imply on the universality of rational co-operation in PDs, and in both "Coordination" and Morals By Agreement, he has been more interested in cases where there are salients, where, by his lights, coordination by a salient is possible. This interest in how coordination can be facilitated by salients when they are available explains, I think, why the possibility that CPs like those we are considering, in the conditions we are imagining, are irresolvable, has received little attention. It is widely assumed that solving perfect information CPs is trivial, and many philosophers tend to advance incautious generalizations which imply the solvability (by pure rationality) of all such CPs. Thus one finds Jean Hampton in her Hobbes and the Social Contract Tradition (Cambridge University Press: Cambridge, 1986) saying of agents facing a CP in choosing a Sovereign that, "nothing is required for the sovereign's institution that [rational]... self-interested people are unable to perform." (p. 172) And we have David Lewis saying in his Convention (Harvard University Press: Cambridge, Massachusetts, 1969):

In considering how to solve coordination problems, I have post-poned the answer that first comes to mind: by agreement. If the agents can communicate...they can ensure a common understanding of their problem by discussing it. They can choose a coordination equilibrium--an arbitrary one, or one especially good for some or all of them....Coordination by means of an agreement is not, of course, an alternative to coordination by means of concordant mutual expectations. Rather, agreement is one means of producing those expectations. (pp. 33-34)

But it is not clear how the agreement can be arrived at in the first place. This problem receives elaboration in our next section. (I suggest above and below that luck is required to solve certain co-ordination problems. Lewis, too, discusses how luck can start

conventions, but he does not invoke it as a possible requirement for making arbitrary face-to-face agreements.)

26. Sobel suggests (in his "Constrained Maximization," pp. 28-29), in another connection, that a rational agent can be imagined to have the power to randomize "in his head," and to commit to acting in a way determined by that process.

27. I thank Nathan Brett, Steven Burns, and especially, Kadri Vihvelin and Sheldon Wein for discussion on these points.

28. My thanks to the reader for this Journal for pressing me on why we can just "insist" in this way.

29. Thomas Hobbes, Leviathan, C.B. MacPherson, ed., (Penguin Books: Harmondsworth, Middlesex, England, 1968), pp. 228-239, 251-257.

30. Hampton, Hobbes and the Social Contract Tradition, pp. 155-160.

31. Kavka too thinks the agents can just "flip a coin." (Kavka Hobbesian Moral and Political Theory, pp. 185-186.) Well sure, if they could agree on which coin to flip, and on who is to make the toss.

32. Hampton, Hobbes and the Social Contract Tradition, pp. 167-168.

33. Hampton, Hobbes and the Social Contract Tradition, p. 169.

34. Hampton, Hobbes and the Social Contract Tradition, pp. 168-172.

35. Hampton, Hobbes and the Social Contract Tradition, p. 172.

36. Hampton, Hobbes and the Social Contract Tradition, p. 172.

37. Or at any rate, such is the part of Hobbes in which Gauthier is primarily interested. Actually, apparently Hobbes wanted the Leviathan not only to enforce compliance with laws (with ideal co-ordinated patterns of choice), but to resolve disagreements about which laws there should be. Communities of agents might not be able to agree on laws because their members might have different stakes in different possible laws, but the Leviathan could choose laws by the standard of which ones would be in His individual self-interest; He would not face a conflictual CP with Himself. (But, contra Hampton and Hobbes, there is no guarantee that He can decide even though He need only cater to His self-interest. For His self-interest requires Him not to provoke revolution, which means He must please His constituency; and if it is divided, any choice of His will displease some of its members, perhaps enough to over-throw him.) However if agents cannot agree on laws, then they cannot agree on a Ruler if that involves agreeing on which kind of Ruler He shall be, i.e., on which laws He will enforce. But suppose the agents,

realizing it is better for all that there be some laws, decide to refer this to a Leviathan; they try to choose a Leviathan not knowing what laws He will favor, or try to choose one whom they know will decide the laws by a randomizer: Now they still face the problem of co-ordinating on which of two possible inscrutable or randomizing Leviathans to choose. So the use of Leviathans may not solve the problem of which laws to have, since we still have the problem of deciding which law-maker to have.

38. Thanks to Sue Sherwin for provoking this line of thought.