

# Social Contract: The Only Game in Town

JAN NARVESON *University of Waterloo*

*ABSTRACT: David Gauthier once said that the social contract offers ‘the only game in town’ if we hope for a rational morality. I argue that he’s correct. Morality consists of rules notionally directed at everyone everywhere. Only individual people are rational, and they have varying interests. The social contract proposes principles that everyone would, in view of their social and environmental circumstances, agree to as constraining their separate pursuits of their ends. There is no other rational way to understand morals, and so it is indeed the only game in town.*

*RÉSUMÉ : David Gauthier a dit un jour que le contrat social constitue la seule option permettant de parvenir à une morale rationnelle. Je soutiens qu’il a raison. La morale se compose de règles notionnelles visant tout le monde, partout. Seuls les individus sont rationnels, et ils ont des intérêts divers. Le contrat social propose des principes auxquels chacun, compte tenu de sa situation sociale et environnementale, consent bien qu’ils limitent les activités par lesquelles chacun poursuit ses buts. Il n’existe aucun autre moyen rationnel de comprendre la morale; c’est donc bien la seule option possible.*

**Keywords:** morals (morality), interaction, social contract, constrained maximization, optimal, game theory

## Introduction

David Gauthier remarks that the social contract offers ‘the only game in town’ if we hope for a rational morality. This paper is intended to support that view, via somewhat more general methods. First, we must define ‘morality’ and identify

*Dialogue* 55 (2016), 695–711.

© Canadian Philosophical Association/Association canadienne de philosophie 2016

doi:10.1017/S001221731600055X



the general context in which it can usefully happen. The idea of morals is that it be a set of all-purpose behavioural directives both internally, directed at the agent him- or herself, and externally, appraising and delivering criticism of the behaviour of others. These directives are, especially, constraints. They are ‘marching orders,’ issued by everybody instead of some central authority, and issued to everybody, in principle. And the background is that we have lots of people who are in contact with each other, actually or potentially, and all at least moderately rational, yet quite strikingly variable in their interests, capabilities, and dispositions—but all, however, vulnerable to the possible depredations of others, and additionally, nearly all both capable of and at least somewhat motivated to engage, if occasion demands, in the sorts of behaviour morality is intended to constrain.

Second, rationality is the maximization of personal utility (explanation needed, and briefly supplied below): people run on their own rational steam. In consequence, if we are to persuade people to accept certain constraints, we must show that somehow morality is in their interests—the interests of individual persons, as adjusted by the facts of social and environmental life—and is so whatever those interests may be. (Or, as discussion proceeds, virtually whatever their interests may be.) Since human interests are variable and often conflicting, it follows that all other approaches to morality are hopeless. What we must do is find commonalities among people that enable us to find the game-theoretically optimal configuration of interests: what is best for each, given that it must be best for the others as well.

I argue that Gauthier’s idea of Constrained Maximization (CM), despite much pummelling in the critical literature, is the right idea about morals, and explain why.

I also argue that Gauthier’s ‘Lockean Proviso’ (LP) actually fills the bill as the fundamental agreed constraint—oddly, not how he seems to see it. Properly understood, it is indeed the only game in town.

## 1. Morals

The title of Gauthier’s by now classic treatise is *Morals by Agreement*, which ought to make it clear enough that his subject is morals. But what’s that? Gauthier is not quite as helpful about that as one might like. A few significant selections help to set the stage.

“But are moral duties rationally grounded? This we shall seek to prove, showing that reason has a practical role related to but transcending individual interest, so that principles of actions that prescribe duties overriding advantage may be rationally justified. We shall defend the traditional conception of morality as a rational constraint on the pursuit of individual interest.”<sup>1</sup>

---

<sup>1</sup> Gauthier, *Morals by Agreement*, 2.

And the phrase “rational constraints on the pursuit of interest” recurs in the next paragraph.<sup>2</sup> Shortly after, he says, “We shall argue that the rational principles for making choices ... include some that constrain the actor pursuing his own interest in an impartial way. These we identify as moral principles.”<sup>3</sup>

In a somewhat later essay, Gauthier says, “... the present claim ... is rather that the principles of justice ... are not principles for rational choice by an individual seeking her good, but principles for rational choice by a society—a group of individuals—seeking justice, and so derivatively principles for choice by each person as a justice-seeking member of the society.”<sup>4</sup>

—and, “‘Ought’-judgements, in the domain of justice, are simply judgments about what is rational for individuals as members of a society to do or to choose.”<sup>5</sup>

—to which he adds, “Of course, I do not claim that this captures all of our ordinary thinking about ethical judgments. Rather, it salvages what is rational in the ragbag of our everyday ethical attitudes. Incorporating the theory of justice into the theory of rational choice is an exercise in rational reconstruction ...”<sup>6</sup>

There is a good deal here that calls for further explanation. Gauthier's ideas have come under an enormous amount of critical consideration, and it is perhaps fair to say that his position has not won wide acceptance, despite wide acquaintance and, I think, admiration. I hope here to add something by way of support for his general outlook. Social contract theory is a view about the foundations of morals. We need to explain what ‘morals’ is, what reason there might be for thinking it has ‘foundations’ and, finally, why it needs to have the kind of foundations he proposes to give it. That reason, as he says, is that the social contract idea is indeed—given certain widely held assumptions—‘the only game in town.’<sup>7</sup> There is no rational alternative.

Our subject is: morals in interaction (or, social morals). It is not, or at least not directly, an essay on the subject of Ethics as understood in Aristotle's celebrated *Nicomachean Ethics*. In that book, generally speaking, Aristotle paints on a very wide canvas: How to Live. He has a theory (of sorts) about that. But only a relatively small part of it is at least ostensibly devoted to the subject we pursue here. Reason may or may not, as Aristotle supposed, be able to supply a canon of precepts about his subject, but only at a few points does he discuss directly the subject here (namely, in Book V). So, what's the difference?

---

<sup>2</sup> Gauthier, *Morals by Agreement*, 2.

<sup>3</sup> Gauthier, *Morals by Agreement*, 2.

<sup>4</sup> Gauthier, “Justice as Social Choice” in *Moral Dealing*, 172.

<sup>5</sup> Gauthier, *Moral Dealing*, 172-173.

<sup>6</sup> Gauthier, *Moral Dealing*, 172-173.

<sup>7</sup> Gauthier used this expression in conversation with Susan Dimock and myself somewhere, sometime, but none of us remembers the details. The phrase itself, however, is remembered clearly.

What makes the aimed-for social contract ‘moral’? We would prefer an answer that doesn’t simply impose an answer by definition; it should hook up with our familiar apparatus of moral ideas, and with the history of the subject. With this in mind, I offer what seem to me to be two useful answers.

## 2. First Useful Answer: Morals as Control of the Passions

The general idea is laid down, classically and correctly, by Aristotle: morals in the most general sense is about the “middle part of the soul”<sup>8</sup>—about actions, but of course actions in relation to passions (taking the term in its broadest sense), which are what impel those actions.

Aristotle’s (and Plato’s) idea is that the passions need constraining. They thought that what constrains those passions is reason. But the trouble is that Aristotle, and later Hume, share the general perception that “reason of itself moves nothing”<sup>9</sup>: in matters of practice, reason follows desire—it’s the slave of the passions. This makes the general conception puzzling.

Clearly, what we must say is that reason controlling the passions is really some passions (or, some mix of passions) being preferred to others—but it takes reason, in considerable part, to help out on this, since things like consequences and close comparisons with alternatives are needed (really needed!). We need to figure out which actions will best satisfy the passions we most want to satisfy, all things considered.

We can agree that introspection on our desires is also welcome, perhaps somewhat contrary to Hume, and even to Aristotle who taught, after all, that our Ultimate End was fixed. That was perhaps a philosophical delusion, or more likely an illusion generated by his terminology. It’s like saying that our Ultimate End is maximization of our Utility—something later theorists do tend to say, even if they don’t quite put it that way. It gives the appearance of a theory or answer, but it’s empty—unless some clear and precise sort of utility is invoked, in which case it’s not empty, but invariably wrong.

## 3. The Mean

Aristotle’s idea was that we need to aim at the mean—an idea he doesn’t make very clear (to put it charitably . . .) Especially, we want to know what makes the middle the middle? Why say that the others are out toward the ‘extremes’? Aristotle doesn’t give anything like a decent answer to that. (He says, in effect, that the extremes are too much or too little—not very helpful!).<sup>10</sup>

But here’s the type of answer Aristotle needs (but doesn’t supply): take eating. Too little—you starve or go into malnutrition and you get anorexic (or whatever) and you die too soon; too much—you get fat and have a stroke, terminating your

<sup>8</sup> Aristotle, *Nicomachean Ethics*, Book I, Ch. 13, and Book II, Ch. 1.

<sup>9</sup> Aristotle, *Nicomachean Ethics*, Book VI, Ch. 2.

<sup>10</sup> Aristotle, *Nicomachean Ethics*, Book II.

life much too soon. The optimum weight is what takes you the farthest in (good) life. Indeed, we can say that maximization of (the good of one's) life calls for moderation in food intake.<sup>11</sup> Other Aristotelian variables succumb to this treatment as well: too much of the variable of which courage is arguably a mean gets you killed, as does too little. (Sometimes, of course, too much or too little gets other people killed, and we then get into the variable with which Gauthier's work is centrally concerned, viz., action in relation to others—morals in the relevantly narrow sense in which Gauthier and we are here concerned with it.)

So, what's the corresponding variable in the social department? That is: which variable is such that we helpfully apply the rubric of 'not too little, not too much' when it comes to dealing with other people? The preferred answer would, we hope, be much the same: maximizing the good of life for the agent—including that of whoever else one desires to maximize the good life of. (Note: unfortunately, we must also include, for the moment, the goal of minimizing the good life for selected enemies, but read on . . . .) Turning our attention to that leaves us with a problem: the good of others is not our good, just as it stands, so how is it that the solution for interpersonal dealings is not simply to exploit them to the maximum? The historically bad solutions to this problem call for the importation of interpersonal utility comparisons. But for the reason just given, we need to do better.<sup>12</sup>

Here's where the social contract theory, *a la* Gauthier (and Peter Danielson and various others), offers us a nice proposal, as applied to our narrower subject of social morals:

1. Too Much is Unconditional Cooperation: The Unconditional Cooperator lets others walk all over him. He gets flattened!
2. Too Little is Straight Maximization: The Straight Maximizer trusts no one, and never reaps the advantages of cooperation. The Straight Maximizer is, to use a useful technical term, a jerk. Ere long, he reaps what he has sown.
3. The Preferred 'Middle': The Constrained Maximizer neither allows himself to be exploited, nor does he exploit anyone else. (In his recent work, Gauthier proposes instead the expression "agreed optimization."<sup>13</sup> But we will stick with the terminology of *Morals by Agreement*.) The CM-er pretty generally reaps the benefits of cooperation, and thus comes out best.

---

<sup>11</sup> cf. Narveson, "Maxificing: Life on a Budget; or, If You Would Maximize, Then Satisfice!"

<sup>12</sup> On a personal note: I was persuaded away from my earlier attachment to Utilitarianism by the strength of Gauthier's "Reason and Maximization," which he presented at the Workshop on Contractarianism at the University of Toronto, 1974 (later published in *Moral Dealing*).

<sup>13</sup> The phrase is introduced in Gauthier 2013.

#### 4. Preface to Social Morals: Separateness of Persons

What makes morals interesting is that it is the social case of ethics—where we are dealing with other people. These other people have their own sets of desires (utilities) and capabilities (powers)—quite variable as they may be. ‘Their own’ means that they are actuated by their desires/capabilities—not mine, or Aristotle’s, or the Pope’s. And therefore, if—as we’d like—we are going to try to lay down the law to those folks, we’d better lay down a law that they have an interest in accepting. Indeed, the fact is that, as Hobbes and Kant and so many others have it, the ‘we’ who ‘lay down’ this law has to somehow include ourselves. For we are not alone, and I benefit only if you comply, while you benefit only if I do. So, the law in question must be such that everybody has an interest in buying into it: I buy it only if you buy it, which you do only if I do ....

In the process, we are driven toward (a) universality, and therefore to (b) impartiality. That is: the law has to be such as to apply to (hence, appeal to) everybody, and since it does, it must be impartial, so that nobody has immediate reason to opt out.

#### 5. People, in General

So, what about these other people, and what about us in relation to them, can make this work? There’s a straightforward, and pretty good, set of answers to both questions.

What, in the first place, do we want from those other people? *Prima facie*, the answer is simple: each wants as much as he/she can get (of whatever it is that one wants).

The social end, for each of us (i.e., from the point of view of A, who is Anybody) may be stated thus: for all persons B, A wants that A’s interactions with B will be maximally beneficial and minimally harmful (= costly) to A, which is to say, in whatever way A counts benefits and harms.

It needs to be emphasized that no interest in or concern for others, as such, is assumed, though it certainly isn’t assumed that we have no such interests. In fact, virtually all of us do, sometimes to a quite dominant degree, motivating self-sacrifice. In short, then, these benefits do not need to be self-regarding, as Gauthier has often pointed out (but not often enough, perhaps.<sup>14</sup>) But it cannot be simply assumed that they aren’t. And especially, the Utilitarian postulate that we all care about everyone else just as much as about ourselves and our particular friends is unacceptable. Conceivably, some few do, but for almost all of us, we don’t.

Even so, we can stick with the formula: we want to maximize what we consider to be benefits relative to costs.

---

<sup>14</sup> In his contribution to this issue of *Dialogue: The Canadian Philosophical Review*, Gauthier provides the most sustained treatment of his conception of human nature (for purposes of contractarian theorizing) to date.

About these other people, there are the Hobbesian characterizations. People are:

- i. rational, which means (originally, anyway) maximizers, of their highly variable sets of tastes, interests, desires, proclivities, using their (highly variable) sets of capabilities (powers).
- ii. of roughly equal 'vulnerability.' This extremely controversial clause refers to people's abilities to make life miserable for others. The bottom line, as Hobbes famously has it, is that "as to strength of Body, the Weakest hath enough to kill the Strongest."<sup>15</sup> (Much more of this, below ...)
- iii. existing in a natural environment of scarcity. This initially motivates the Hobbesian motive of competition. But the scarcity in question is relievable, especially—and open-endedly—by cooperative effort: successful cooperation improves life for all the cooperators; its absence gets us a miserable life and an early demise, as Hobbes even more famously put it.
- iv. pretty generally, if pretty variably, non-altruistic (and, often, downright egoistic).
- v. amoral: people are not presumed to be by nature actuated by any moral input.

These are what later writers, notably Rawls and Gauthier, call the "circumstances of justice," or more generally, as I would want to add, the 'circumstances of (the social portion of) morals.' Hence, our subject.<sup>16</sup> They are the circumstances of justice in particular because, as Hume points out, without these features, we either could not have or would not need a virtue of that kind. But we do have them, and so we must go from there.

An important and disquieting possibility is that some will have natively malevolent interests. From our historical perspective, this possibility is underrated by classical authors. Hobbes seems never to contemplate the possibility that people might act from sheer hatred, rather than from "competition, diffidence, or glory." We must, of course, confront it. But it is of the nature of such possibilities that they may be tweaked up by the ingenious philosopher, and put well beyond the reach of any sort of reason—sci-fi movies attest well enough to that. The social contract view, however, correctly settles for the only option: we are at war with such people, and let's hope we'll win, which we likely will, there being so many more of us. CM allows unlimited maximization against unlimited maximizers, even if what they maximize is such singularly unproductive magnitudes as hatred or morbidity.

---

<sup>15</sup> Hobbes: *Leviathan*, Ch. XIII.

<sup>16</sup> See "Reason and Maximization," 137 for Gauthier's summary of Hume's account. Note especially his (5) and (6), which slightly elaborate on the 'vulnerability' in (ii) above.

## 6. Second Useful Answer: Social Behaviour Controls

We first looked at morality as the control of the passions, the idea being that the control was essentially internal. We turn now to a related but still distinct sense of the term: morality is a set of decentralized (informal) social behaviour controls. That is: We (all of us) use our (very variable) resources to influence other people's behaviour so as to promote the achievement of our ends. At least most people are quite reachable by methods of this general type: the mother's influence via displays of affection or disapproval, people's susceptibility to what their peers cheer or boo for, and so on.

But morals is about this control by all of us, and, of course, almost all of us are other people. My solo influence is trivial. Our united influence is great. So, the question is whether there is a reasonable prospect that the influence of many will be in the right direction.

Of course, this is a question that can be taken empirically as well as abstractly. And at the empirical level, alas, there is plenty of variability among moral codes around the world and through history. Even so, there are also some commonalities. Much as we differ, we all have interests in life, health, and non-invasion. The core of morality is a prohibition on murder, more generally, uninhibited harming of others, and being unreliable in one's practical commitments. But most moralities also have assorted more apparently arbitrary prohibitions and requirements. Theorists about morals think they can do better. The abstract, game-theoretic approach, we think, has the potential to zero in on commonalities in a profitable way, and to sift out the bits of chaff found in too many local sets of mores and morals.<sup>17</sup>

## 7. Why Morals? Good and Bad Answers

Why should we have morals at all? The good answer is: because we'd like to be able to have the best means for managing general encounters with fellow humans—especially, people not members of our family, friends, or close associates, all of whom we can generally trust. But morality is concerned with all, and therefore with the by far most who are not in the latter classes, even though their doings can be and often are crucially important to each of us. We'd like this because we are in society—we bump into each other, often. We'd like to know what we can expect from them, and vice versa.

The history of philosophy is amply supplied with bad answers to this question of the foundation of morals—numerous, and some worse than others—such as because:

---

<sup>17</sup> Note on politics (perhaps necessary for some): in this inquiry, we are not presupposing, or operating in, an environment of political institutions. Politics is logically downstream from morals: we criticize political happenings on the basis of moral considerations. We don't suppose that political institutions have independent authority. Hence, nothing is said about them here, large though they loom in typical lives, the world over.



- God tells us to behave some way or other,
- Nature tells us to behave some way or other,
- Intuition tells us to behave some way or other,
- Pure Reason tell us to behave some way or other,
- most of our neighbours tell us to behave some way or other.

## **8. What Makes the Bad Answers So Bad? Reflections on the Thirty Years' War**

What the bad answers have in common is that they are philosophical non-answers dressed up as answers. Take, for example, the tendency to invoke God. Once upon a time, there was a different kind of 'morals by agreement': everybody agreed that Roman Catholicism was the Truth About It All—until somebody smelled a rat. Then, they had the Thirty Years' war. After reducing the GNP in that part of the world by about 75%, and the population by maybe 30% (70%, in some regions), the (remaining) Catholics were still Catholic and the (remaining) Lutherans were still Lutherans, and it began to dawn on people that maybe this is, shall we say, suboptimal.

So, religious liberty is the way to go, folks. Why? For one thing, because there is no way to persuade people, using ordinary ('public') reason that one party is in the right about religion, all others being in the wrong. It's fairly easy to just let everybody practice their religion in peace, so long as it really is peace (that is, attempts at conversion are strictly by voluntary means.) A modest amount of Aristotelian control over the religious passions should do the job, with most normal people. (Alas, as we know, it doesn't seem to work with a fringe—'jihadists' and other religious fanatics are still among us.)

And, for another, the religious story about morals is conceptually hopeless. If some supposed supermind tells us that this or that is the right thing to do, and we pause to inquire why she or he would think so, what answer can that personage supply? 'Well, I just give the orders around here, no back-talk being tolerated' is not a very satisfactory reply: God had better do better than second-rate nannies. But, if he does reply, what's He going to say, if not that following these or those rules will be best for us, considering how we are and how we relate to others. But that story stands on its own: divine middlemanships add nothing, though they do tend, historically, to subtract a lot, with unhappy results for humankind.

Nature? That might be a name for our project: that is, we go from the way people are, to assembling good rules for the group. But most Natural Law types have the delusion that we can read the rules off the trees and such, without doing our homework in the way of decision-theoretic thinking about interaction. But nature, of course, doesn't say anything, hence doesn't tell us anything—it is we who have to work out the answers.

As to moral Relativism—insofar as the morals of Group 1 conflict with those of Group 2, what happens when persons from G1 bump into persons from G2? 'My way is better because it's mine!' is the very form of a non-answer .... And,

that the rest of us are necessarily right about what I, or Jones, should do is, shall we say, decidedly unobvious.

Gauthier refers to a “foundational crisis”<sup>18</sup> in moral theory. This stems from those failures—the attempt to ascribe purpose to the world at large, or to impute some sort of pre-emptive authority to somebody other than the very agent whose actions are in question here. That way lies emptiness, and eternal dispute, which is pretty much the same thing.

## 9. Rational Morals

The project is to find a source of rational, unified general control, administered by all, ‘legislated’ by each, capable of providing effective guidance. Plenty of *de facto* sets of morals are considerably less than rational, as noted. What we philosophers seek is the most rationally endorsable such set—the set of principles supported by reason.

Why ‘unified’? Because we’re dealing with everyone here. If we say one thing to A, and another thing to B, A, or B (probably, whoever gets the worse of it, and because one of them just might be a philosopher) will want to know: why should I accept this? If we lack a good answer, then morals lacks reason.

## 10. The Social Contract: Why?

As Gauthier says: there is only one way to go, given the aforementioned crisis (crises, maybe): we must make up our morals from our separate interests. There is nothing else. That’s why it’s the only game in town.

The only respectable answer to the question, ‘Why?’, is this: given that we’re among other people, who are different, but capable of affecting me seriously for better or worse, this is the best set of rules for me to endorse as proposed rules for the whole group. And remember that ‘me’ is each person there is—not just the author of this piece, or the reader.

It is important that there is no other rational alternative. There is no point in trying to lord it over the rest, and certainly no point in being a patsy. Proposing to run them under the juggernaut of intuition or such is useless. Morality can’t be a black box. If the rules don’t make sense to the people they are supposed to govern, they will fail as controls. And, if they do that, then they’re pointless.

## 11. Why CM?

CM is the disposition to cooperate with cooperators, and not to do so with non-cooperators. Why is CM the way to go? We want it—if we can have it—because, without it, we are in the Hobbesian rule-free condition. And I (along with Gauthier) broadly accept Hobbes’ thesis: absent any rules, and given the way we are, we are in for a condition that makes life “solitary, poore, nasty, brutish, and short.”<sup>19</sup> More precisely, lacking mutual cooperation, we are

<sup>18</sup> Gauthier’s term: see his “Why Contractarianism?,” 15.

<sup>19</sup> Hobbes: *Leviathan*, Ch. XIII, 104.

headed for suboptimality—situations in which some could be better off without others being worse off. And, if we run this back far enough, we would indeed arrive in the Hobbesian condition: much worse for all.

While I broadly accept Gauthier's account of CM, I think we should dissent from what we may call his 'Transcendental Deduction' of it. Suppose that rational men start out as 'straight' maximizers: they will, upon contemplating the costs of non-cooperation, Gauthier says, change their theory of rationality. CM will become the criterion of practical reason. It is difficult to follow the metaphysics of this: is our "theory" of rationality something we can intentionally, voluntarily change? That's hardly obvious.

Paul Viminiz delightfully refers to what he takes to be Gauthier's own solution to this as the "pharmaceutical" way: we take a pill, which turns us into CM-ers, and we carry on.<sup>20</sup> But of course there is no such pill, literally. What do we say instead, then? The first thing to do is recall Aristotle: ethics is about the control of the passions in determining action. Intellect of itself, says Aristotle, "moves nothing."<sup>21</sup> If we say that the man who looks before he leaps, or who disposes himself to look for signs of cooperation among his interactees, acts rationally, what makes that true is his attentiveness to overall advantage. Recall too that our overall advantage is a matter of realizing our preferences, and those preferences are not essentially 'rational'—they are in the 'middle' part of the soul. People are capable of being directed by reasoning, of course, but still, what are so directed are still desires, passions, and dispositions to act. So, when we ask whether we would do better to be constrained in our maximization, we are indeed, as Gauthier says, asking whether we should cultivate a certain disposition or habit. What is rational is to acquire, and develop, this disposition in view of its expectable returns for the good life, in the long run and, always, in society.

It is customarily argued that defection in the prisoner's dilemma (PD) is the rational strategy because it is the dominant strategy. But, if the players have the option to cooperate, as they by definition do, and if they can communicate with each other, which they do not do just by definition, then the prospects of gaining by defection usually ranges from low to zero. That we would defect against people we don't trust, provided we aren't going to engage in indefinitely iterated play with them, is no doubt true. Is trust rational? It is a disposition, in the middle part of the soul—so, of course, it is not by definition rational (or irrational). Yet, it is rational to develop trust in relation to people who can do likewise.

Moral philosophers who want to make out that moral behaviour is rational have a problem: man by definition is 'the rational animal.' So, if morality is

---

<sup>20</sup> Viminiz has discussed the "pharmaceutical" route to becoming a CM-er over many years and in countless conference presentations, including in Viminiz 2013.

<sup>21</sup> Aristotle, *Nicomachean Ethics*, Book VI.

rational, in the sense that moral behaviour is entailed by rationality, then it follows that all men are moral. But they aren't, always! And it is indeed among us imperfectly rational beings that morality is both (a) an institution (unlike rationality), and (b) an extremely important one.

Gauthier famously denies that defection is rational even in one-shot PD games. But, in his exposition of CM, he points out that we need evidence that our interactees are also CM-ers before we can risk cooperation. And how do we get this evidence, if not from past performance?

Social interaction reinforces CM—and, of course, does so very strongly. And it does, of course, because it needs to. A moral theory purporting to dispense with this aspect of the morality we know has, one must say, gone off the rails.

## 12. Mechanics of CM

We said: 'The CM-er pretty generally reaps the benefits of cooperation, and thus comes out best.' Well—at least, he does if he manages to cooperate only (mostly?) with other cooperators. Or, more precisely: he does better insofar as he deals with fellow cooperators.

The would-be CM-er needs two things:

- a) to be able to detect the cooperation trait in others, and
- b) to be able, and ready, to cooperate even if he could gain by cheating.

Both of these are extremely thorny matters, in principle. A huge amount of the literature is devoted to worrying about them. And yet, the funny thing is that they aren't all that much of a problem to most of us most of the time. We pretty routinely deal with people along lines of trust and fruitful cooperation; we very frequently refrain from cheating even though, if we just thought about it a bit, we know that we could get away with it. But it scarcely occurs to us even to try, and even then most likely it is analytical philosophers abstractly addressing the possibility, even as they ignore the goodies on the shelf, pay their bills reliably, and walk off unperplexed.

I wonder why? That's an interesting fact about people. It might be taken as some empirical support for the contractarian position. It's certainly empirical support for the proposition that people are social beings, susceptible to acculturation. Unfortunately, some have taken it as evidence for Natural Law, or—if this is any different—Intuition. But both of those have the same problem. In what sense is a law 'natural' if we can break it and often do? And of what use is intuition if we have no idea how it works, what its reasoning is?

So one interesting question is: if this is such a thorny theoretical issue, why is it so frequently not a problem in real life? We can be sure that the answer is: because of very extensive, very widespread, very frequent iteration, among persons one comes to know and, in consequence, trust.

Is this like taking a pill? If we call it that, it's a pill of our own making—a pill of the spirit, as it were, rather than a material one. And, to be sure, it is not

entirely of 'our own' making: there are parents and peers and such too help out. Yet, they can only do so much. What's in our souls is what ultimately matters here.

Nevertheless, it seems that almost all of us are pretty good at popping these things. And that, I suggest, is really enough. After all, there is no way to have 'more' without running afoul of the facts of life.

At this point, let's remember what our project is. We want to know whether any among the possible sets of principles, rules, 'directives' is suitable for universal use: whether we can expect any such set to be conducive to the best life for each person, if others adhere to it as well. That's the question posed, and answered, by the social contract.

### **13. The 'Contract with Everybody'**

Contractarianism is generally represented as a sort of 'agreement among everybody.' In just what sense is it such an 'agreement'?

That is an important question. Obviously, it is not any sort of historical agreement with 'everybody'—that being obviously impossible. And too, alas, just about no matter how you slice it, it surely appears that some people either just never 'made' it, or somehow fail to pay much attention to it, quite a bit of the time. So, how can we claim to have this 'agreement'?

Well, in the first place, we don't, quite. What is claimed is only that it would be rational for everyone to 'sign on' to this. The gimmick here is that the terms of the 'contract' are the common good, in the specifically liberal form of the maximum benefit from interaction compatible with similarly maximal benefit for all. Thus, as emphasized in the previous pages, we all have a stake in this.

In the second place, there's another aspect to this: suppose the proposal I make to everybody is: 'Here, I offer cooperation. If you cooperate, I will too. If you won't, then we're in conflict and I act accordingly. What do you say?'

Well, it turns out that this is an offer you can't refuse! The contract is universal because all bases are covered: either we do agree to cooperate or we don't, in which case we have 'disagreed.' (Do we say of someone who breaks an agreement he signed that he didn't really sign it?) Those remaining in the State of Nature—the unagreed state—have presumably agreed in the second sense, the rest of us in the first. They somehow prefer war to peace, it seems, and we, the rest of us, are in the situation of all State of Nature dwellers in the sense that regarding them, all bets are off. They extend us no rights, so of course we extend them none either. The fact that that is then their situation is precisely what, one would hope, would attract them to the social contract in the preferred sense in which it is an agreement to 'lay down our arms'—to confine our relations with our fellows to the peaceable ones.

So, what's the problem to which CM is a solution? The answer, I think, is: will the Rational Person constrain himself, if he should have the opportunity not to do so, but gain by it? That's Hobbes' question re the Foole, and Hume's re the Sensible Knave. The Rational Person will certainly constrain himself if

he believes that such constraint is both necessary and sufficient for getting the benefits of mutually cooperative behaviour. But, of course, it is neither, in the large picture—not, that is, on each particular occasion. Some people do gain by reneging, and some do lose by ill-advised cooperation.

So, were the losers irrational? Or were they just unlucky? Indeed, were the winners irrational? (If perhaps lucky.) Like Socrates so long ago, Gauthier maintains that even the winners were indeed irrational. In what sense? Presumably that they played a strategy which depended for its success on fooling people (in Hobbes' words—invoked by Gauthier—they succeed only “by the error of them that receive him”<sup>22</sup>).

Of course, the next question is why a rational person wouldn't fool people if he could. There is a moderately plausible claim that you can't fool all the people all the time, true. But perhaps fooling enough of them enough of the time is good enough—why not?

The more such people there are, of course, the more resources other people must expend in trying to track them down and deal with them when found; and, the more people are tempted to go crooked instead of straight. We could, no doubt, make estimates of the marginal payoff to doing the one or the other, and recognize that our punishments need to be adjusted so as to maximize the incentive to become faithful CM-ers.

But there's a good deal more to it. Read on!

#### **14. The Other Aspect of Morals**

Of course, our question is about morals. Is there a distinction between a rational morals and a rational line of personal behaviour? Most people think so, really. So, how does this work?

One good answer, I think, lies in thinking through the fact that morals is a matter of social reinforcement. So, in relation to any individual, it has two aspects:

- a) complying (or not) with its rules, and
- b) reinforcing/publicizing/inculcating those rules.

If we're asking, What does reason tell us to do?, then regarding (a) it might council immoral behaviour, at least on occasion. But then there's (b)—what should we say in public? That is, how should we behave in the matter of scrutinizing and appraising people's behaviour (others' as well as our own)? What should we do in the way of publicly supporting and condemning various lines of behaviour?

There's now room for a new suggestion: you'd have to be an idiot to come out publicly in favour of war, confusion, hoodwinkery, and the rest of it.

---

<sup>22</sup> Hobbes: *Leviathan*, Ch. XV, 121.

People who do this usually dress it in sheep's clothing: these actions that appear to be so evil are, they will say, really good, contrary to appearances .... (Persons in positions of political power are especially likely to take this line. They're virtually guaranteed to, actually!) Thus we can expect, and understand, a lot of slippage twixt cup and lip. Hypocrisy happens—and not just among politicians. The point is, though, that the compulsion to say such things is very strong. For otherwise, you might as well put a sign up: 'Don't bother to deal with me!' And since practically everything we have comes from others, that is a death warrant, dialectically speaking.

### 15. The Status of the Lockean Proviso: A Proposed Revision

Gauthier presents his three main theses in this order:

- First: the bargaining solution, Minimax Relative Concession (MRC)<sup>23</sup> (now modified to Maximin Relative Benefit),
- Second: Constrained Maximization (CM)<sup>24</sup> (now modified to Agreed Optimization—perhaps a better name for it ...),
- Third: the Lockean proviso (LP), which forbids forwarding one's own benefit by imposing detriments on others—bettering one's own situation by worsening that of others.

But I hold that these are in the wrong order.

First should come CM,<sup>25</sup> indeed: the disposition to cooperate when cooperation is possible is the basic device for getting us out of the State of Nature.

But the immediate output of CM, I argue, is the LP—not MRC: principles for dividing shares of 'social product' are strictly and completely subordinate to the LP. First we constrain individuals from pursuing their utility by imposing disutility on others. Then we bargain, in individual cases, that is, productive interactions regarding particular products or services. The LP fundamentally determines (or, enables us to determine) what is whose; and further division of product is properly done by the market—not by a further moral rule. In the market, after all, everything is done by agreement.<sup>26</sup> If we are serious about 'morals by agreement,' the market is the way to go. Yes, there are complex stories about public goods, and, indeed, morality itself is fundamentally a solution to a public goods problem: the tendency to get what we want by

---

<sup>23</sup> See Gauthier, *Morals by Agreement*, 140 (more generally, Ch. V, 113-156).

<sup>24</sup> See Gauthier, *Morals by Agreement*, 167-170 and more generally Ch. VI, 157-189.

<sup>25</sup> See Gauthier, *Morals by Agreement*, 205, and more generally Ch. VII, 190-232.

<sup>26</sup> Note: The term 'market' is used here to refer to such social exchanges and institutions for accommodating it that really match the defining characteristics of markets—not to (what's left of) the 'market' as hyper-regulated and distorted by well-meaning governments.

taking it from others, without their consent, is our problem. Adopting an internalized aversion to doing things that way is what is needed, and is what morality basically consists of.

The LP prescribes that one's actions should be (at least weakly) Pareto superior to the status quo. We are always trying to improve our own situations (and we often succeed). But we are not to do so by 'worsening the situations' of others. Commercial activity improves the producer's situation by improving the consumer's—this is win-win.

The LP has a distinguished pedigree. It is the rule of Peace—Hobbes' first Law of Nature; it is Locke's Law of Nature; it is Kant's Universal principle of Justice; it is Mill's Principle of Liberty. And it might be Rawls' First Principle of Justice.<sup>27</sup> It is ubiquitous in moral codes. All this is no surprise.

Hobbes declares that all the other Laws of Nature are derived from that first one. I claim that he's right, though many disagree or think they do. If he is, then once we have the basic argument for the first law, everything else follows. And that basic argument, I have held, is provided by Gauthier. The intelligent individual adopts CM—along with all others. He does so because it supplies the foundations for the only general moral principle that can be agreed to, in their own various interests, by all (or, all but the totally incorrigible).

The Beatles said: 'All You Need is Love.' They were wrong. All we need is Peace! Peace is the fundamental public good. And it has the usual public goods problems. I can only get it from you, and you from me—there's no way to confine its benefits to the producer. And temptations to cheat are ubiquitous. So, morals is an uphill struggle.

But it's a struggle worth engaging in, because the social contract, as understood here but basically Gauthier's, is indeed, the only game in town—where, the town is the domain of rational moral theories.

---

<sup>27</sup> In its original form, Rawls' First Principle says that everyone is entitled to the maximum liberty compatible with the like liberty for all. After that, it gets ramified in odd ways. I have explained, in "A Puzzle about Economic Justice in Rawls' Theory," why we might as well skip the supposed Second Principle.

## References

- Aristotle  
 1912 *Nicomachean Ethics*, Ross translation. Oxford: Clarendon Press.
- Gauthier, David  
 1975 "Reason and Maximization," *Canadian Journal of Philosophy* 4 (3): 411–433.
- Gauthier, David  
 1984 "Justice as Social Choice," in *Morality, Reason and Truth*, edited by David Copp, 251–269. Totowa: Rowman & Allanheld.



- Gauthier, David  
1986 *Morals by Agreement*, NY: Oxford University Press.
- Gauthier, David  
1990 *Moral Dealing: Contract, Ethics, and Reason*, NY: Cornell University Press.
- Gauthier, David  
1991 "Why Contractarianism?," in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 15–30. Cambridge: Cambridge University Press.
- Gauthier, David  
2013 "Twenty-Five On" *Ethics* 123 (4): 601–24.
- Hobbes, Thomas  
1651/1950 *Leviathan*, NY: Everyman.
- Narveson, Jan  
1976 "A Puzzle about Economic Justice in Rawls' Theory," *Social Theory and Practice* 4 (1): 1–27.
- Narveson, Jan  
1978 "Rawls on Equal Distribution of Wealth," *Philosophia* 7 (2): 281–292.
- Narveson, Jan  
2004 "Maxifizing: Life on a Budget; or, If You Would Maximize, Then Satisfice!," in *Satisficing and Maximizing: Moral Theorists on Practical Reason*, edited by Michael Byron, 59–70. Cambridge: Cambridge University Press.
- Viminitz, Paul  
2013 "Morals by Agreement versus Morals by Treaty," a conference paper delivered at the conference celebrating the 25<sup>th</sup> anniversary of publication of *Morals by Agreement* at York University (May 2013).