

# Getting the Baseline Right—or—Why I’m Right and Everyone Else is Wrong, in each of the Two Senses of ‘Why’

PAUL VIMINITZ *University of Lethbridge*

*ABSTRACT:* My fellow contractarians and I are of a mind that it would be irrational to comply with a distribution of the cooperative dividend that worsens one’s condition. But worse than what? According to David Gauthier et al., it’s non-interaction, i.e., what would be the case were the negotiators never to have met. I argue that it’s what would be the case in the absence of their coming to an agreement. As it turns out, this distinction can be, and often is, a matter of life and death.

*RÉSUMÉ :* Mes collègues partisans du contractarianisme et moi sommes d’avis qu’il serait irrationnel de se soumettre à une distribution de la dividendecoopérative qui empirerait sa propre condition. Mais par rapport à quoi peut-on dire que cette condition serait «pire»? Selon David Gauthier et al., elle serait pire que la non-interaction, c’est-à-dire ce qui se produirait si les négociateurs ne s’étaient jamais rencontrés. Je soutiens plutôt qu’elle serait pire que le cas où ils ne seraient pas parvenus à une entente. Il se trouve que cette distinction peut être, et est souvent, une question de vie ou de mort.

**Keywords:** baselining, contractarianism, Pareto, worsening, Hobbes, Locke, Gauthier, self-effacement

*Dialogue* 55 (2016), 739–757.

© Canadian Philosophical Association/Association canadienne de philosophie 2016

doi:10.1017/S0012217316000676

## 1. Introduction

### 1.1. *Our Common Project*

Here, I take it, is our common project:

We are, quite literally, civil servants. We work for the polity that pays our salaries. Our job is helping our clients get their moral and political reasoning right. But there's no fear of featherbedding here. There's a significant divergence between what moral and political arrangements our clients currently enjoy and those that would be sanctioned by more careful reasoning.

Some of us—Jan Narveson, for example—think this divergence is a veritable disconnect.<sup>1</sup> And so he advocates a radical revision to those arrangements. Others think the status quo requires only fine-tuning. But if any of us thought the way things are is exactly the way they should be—and can be pretty much counted on to remain so—she'd be hard-pressed to say what work her moral and political philosophizing is doing.<sup>2</sup>

### 1.2. *Prescriptivity, Descriptivity, and Justification*

But, if we're in the business of saying how things should be, of what relevance is the way things are? Twofold:

First, things can only be the best they can be under a given set of circumstances. For example, according to Hobbes it's that "two men desire the same thing, which nevertheless they cannot both enjoy"<sup>3</sup> that makes morality and politics necessary. But, he adds, what makes morality and politics possible is that many of the commodities over which we compete, though scarce, can be shared.<sup>4</sup> In other words, we're facing only moderate scarcity. Game theoreticians capture this by saying we're faced with (what they call) 'mixed-motive games.' Or, in yet other game theoretic terms, by the 'moderate scarcity condition' is meant the availability of a cooperative dividend.

But though morality and politics, coupled with science, should, properly deployed, increase our productivity, we're never going to produce such plenty that we're going to completely do away with scarcity, and so with the need for

---

<sup>1</sup> See Narveson's entire corpus since and including *The Libertarian Idea*.

<sup>2</sup> That a normative theory must not exactly coincide with the way we're currently doing things can be contrasted with two related but different principles: John Rawls' (wide) reflective equilibrium and Steven P. Lee's Principle of Tolerable Divergence. Rawls suggests a dialectic between our theory and our considered moral judgments. The Principle of Tolerable Divergence insists that "the moral norms an institution prescribes [must] not greatly diverge from what [its] prudential norms prescribe." (pp. 21, 22, 27, 273, 308, 309.)

<sup>3</sup> Hobbes, 184.

<sup>4</sup> Not so, of course, under conditions of famine, which is why moral dispositions and political arrangements break down under such conditions.

morality and politics.<sup>5</sup> Neither are we going to completely do away with competition, and hence the need for morality and politics, by learning to content ourselves with less.<sup>6</sup> We can reduce competition, from which, says Hobbes, "proceedeth diffidence ... thence Warre," thence the need for morality and politics.<sup>7</sup> After all, a man who wants nothing wants for nothing. But Hobbes knows of no such man. And neither do I.<sup>8</sup>

Can we imagine a circumstance under which humans aren't faced with mixed-motive games? Certainly. Total war and the Hereafter, to name but two.<sup>9</sup> But *Leviathan* isn't addressed to total warriors or angels. It's addressed to us. This is why Chapters I through XIII offer a description of the human condition as Hobbes finds it. Only in Chapter XIV does he then turn to how to make the best of our bad situation.

Are there, as Virginia Held has argued, human communities and/or times to which Hobbes' description does not apply? Certainly. Hobbes himself acknowledges this when he refers to "the government of small Families, the concord whereof dependeth on naturall lust."<sup>10</sup> By this Hobbes doesn't mean that families don't face scarcity. Rather he means that in a family—indeed what we mean by family—is that, though "two [members may] desire the same thing which neverthesse they cannot both enjoy," neither has a desire to consume these scarce resources at the expense of the other.<sup>11</sup> That said, scholars who

---

<sup>5</sup> Pace Marx, who thought that there would come a time when the dictatorship of the proletariat could eventually do away with itself.

<sup>6</sup> Pace Buddhism and the environmental movement.

<sup>7</sup> Hobbes, 185.

<sup>8</sup> There's a theory of the good according to which the goodness of states of affairs varies directly with the satisfaction of desires, from which it follows that the best world is one in which we desire nothing, since then all our desires are satisfied, albeit only trivially. An alternative theory is that the good varies with the number of desires satisfied, from which it follows we should multiply our desires. A third view is that good varies with the ratio of satisfactions over frustrations. And so on. My own view is that all these views are just silly. Nothing is of value save to a valuer. Accordingly, a world in which many valuers are satisfied is better than a world in which fewer are satisfied only to a valuer who is more satisfied by more valuers being satisfied. To think otherwise is to commit—as did J.S. Mill—the fallacy of composition.

<sup>9</sup> Most wars are characterized by the following: much as I might want to displace you from your hillside, not if all I'm going to be able to plant there is my tombstone. So war is a mixed-motive game. Hence the rules of war. The objective of total war, by contrast, is the extermination of the enemy, even at the cost of one's own demise. Hence no cooperative dividend, hence no mixed-motivity, hence all rules barred.

<sup>10</sup> Hobbes, 187.

<sup>11</sup> Opponents of same sex marriage/advocates of family values would do well to at least consider this definition of the family. Hobbes, 184.

think that Hobbes is a psychological egoist interpret him here as saying that it's only mutual exploitation that holds the family together. And this seems clearly false, since family members typically care for each other even when there's no prospect of further exploitation. Others, myself included, hold that Hobbes is not a psychological egoist. By our lights, all he's saying here is that small groups of people can sometimes govern themselves by the bonds of affection only, however that affection is motivated.

What's in dispute, then, between hardcore contractarians and their detractors, is whether governance driven not by mixed-motivity but by bonds of love is part of morality. Our detractors say the promotion and/or maintenance of such love is a matter of morality. Indeed some of them argue that love is the genesis of morality. Contractarians, by contrast, say love—or, for that matter, boundless hate—is a 'morality free zone.'<sup>12</sup>

But we also say that nothing substantive can hang on this dispute. For to say that morality arises only out of mixed-motivity is not to say that interactions that aren't mixed-motivated don't warrant the attention of philosophers. It's to say only that they don't warrant our attention qua moral philosophers.

The two camps are in some measure synthesized by Hume, according to whom moral dispositions, however motivated, can't be activated without the sentiment of fellow-feeling. Hume may well be right. But contractarians give away nothing in conceding this. We simply say that

- a) the adoption of moral dispositions required to maximize under conditions of mixed-motivity and,
- b) the extension of familial fellow-feeling to non-family members,

may be phenomenologically indistinguishable.

But if the phenomenology of the moral point of view doesn't distinguish between the Hobbesian account and the Humean—and since the etiology of morality is shrouded in our evolutionary history—why should we prefer the Hobbesian account to the Humean? Worse yet, suppose there were good reasons to believe—and evolutionary ethicists have shown there are such reasons—that morality owes its perdurance in us not to individual selection but to group selection. Suppose, in other words, that morality isn't, as Hobbes would have it, a strategy by which one maximizes his own fitness under conditions of mixed-motivity, but rather a strategy for maximizing group fitness. Shouldn't we, then, prefer the Humean account to the Hobbesian?

To which we answer that, phenomenological under-determination notwithstanding—for that matter historicity notwithstanding—only the

---

<sup>12</sup> Contemporary contractarians owe this pithy but highly effective phrase, as well as many others, to David Gauthier.

Hobbesian can speak to the Foole. That is, only the Hobbesian can convince the psychopath why, notwithstanding his being bereft of any morality, he should, if he can, find himself some morality with all possible dispatch. For the central question for contractarians is not how we came to have the moral dispositions we have—and/or the political arrangements we currently enjoy—but rather what dispositions and/or arrangements ought we to adopt now. To adopt whatever dispositions and/or arrangements that account for our being here to ask the question would be to commit the naturalistic fallacy. Yes, I probably do owe my very being to group selection processes. But what is that to me? So contractarianism—and only contractarianism—offers its clients a justification for morality, i.e., a justification for morality from non-moral premises, and a justification pegged to our clients' occurrent desire-sets. Moreover, contractarianism can tell our clients which moral dispositions (and political arrangements) to adopt (or promote), given those occurrent desire-sets.

### *1.3. Bounded Rationality*

The second way normative philosophy involves attention to the way things are is that one can't get to where she should be without knowing where she is. For example, human rationality is bounded rationality. So sometimes it's not enough just to say to our clients, 'Look, here's a better way!' For looking is not necessarily seeing. Sometimes our clients' bounded rationality makes it impossible for them to see.

But, as Plato pointed out, that doesn't relieve us of our professional responsibilities. For example, I've argued in defence of terrorism. It was my professional responsibility to do so. But that doesn't mean I should press that argument at a memorial service for the victims of 9/11.

That said, just as the Pope takes pains to make clear when he's speaking ex cathedra, we too must make clear when we're speaking to the hoi polloi in the cave and when we're speaking to each other outside of it. And indeed, in this paper, I'm speaking not to our clients but to my fellow philosophers. So, here at least, no mincer of words need I be.

Another example of how bounded rationality informs our advice is that getting from here to there may have to be done incrementally. For example, we could save billions of dollars in the cost of air travel by doing away with the practice of farewelling, or at least moving it from plane-side to the front door of the house. But, though it would be optimal to make the shift in one fell swoop, people are too wedded to the practice for the move not to be done in increments. In fact, not only would nothing be gained by suggesting such a radical revision to the practice, such a suggestion might even provoke a return to plane-side farewelling as a protest. Weather concerns allowed us to move it from plane-side to the gate. Then security concerns moved it another hundred yards back. The cost of parking will gradually drive it from the airport to a mall near the airport. And, finally, highway congestion

and the cost of gasoline will drive it to the local mall. So we philosopher-kings will get our way. We just have to be patient.

#### *1.4. Error Theories and their Probative Force*

One can get from where-she-ought-not-to-be-but-is to where she should be without knowing why she's where-she-ought-not-to-be-but-is, i.e., what mistake she might have made in virtue of which she's where-she-ought-not-to-be-but-is. (What philosophers call) an error theory isn't a necessary component of a complete political theory. Nevertheless it's a highly useful check. That is, that I'm the only person on the planet to think that *p* is no reason to think I'm wrong. But it would be a reason to at least double-check my reasoning if I couldn't come up with an account of why everyone else believes, albeit falsely, that  $\sim p$ .

This is important. Suppose someone believes, albeit falsely, that *p*, and she believes this because she believes—only this time rightly—that *q* and that *p* implies *q*. In other words, she believes that *p* because she's affirmed the consequent. Affirming the consequent is a common mistake in human reasoning. But there's a perfectly adequate explanation as to why it's so common. It's so common because most conditionals, at least in the domains in which the error is so common, are also bi-conditionals, and so inferring the other atomic from the atomic given as a premise is, for those domains at least, a heuristic with a success rate adequate to compensate for its occasional failures. But this is not the case with, say, a complete non-sequitur, e.g., *p* therefore *q*. So if someone believes *p* and concludes *q*, chances are it's because she believes *p* implies *q*, or some other suppressed premise.

So what's the cash value of all this? Just that, all other things being equal, my political theory is better than your political theory if, by supposing people's errors arise from an otherwise laudable heuristic, I can account for the divergence between my theory and how people actually behave, whereas to account for the divergence between your theory and how people actually behave, you need to say (something like) 'people are just stupid.'

This is not to say (something like) 'people are just stupid' couldn't be the case. It's to say only that 'that people are just stupid' is not as good an error theory as that they're employing an otherwise laudable heuristic. For if people were stupid in this domain, why aren't they equally stupid in other domains? But if they were equally stupid in all domains they'd have long since gone extinct. So people-are-just-stupid is a bad error theory. Of course, having a good error theory isn't a knockdown proof of the theory it accompanies. But, as I say, having a bad error theory is a reason to at least double-check it.

What makes one error theory better than another is no different from what makes any theory better than another, and this includes which theory has the better second-order error theory. And so on.

Accordingly, what I want to do in what follows is threefold. Though I've advanced this argument in greater detail elsewhere<sup>13</sup>—I want to rehearse the case that

- 1) I'm right and everyone else is wrong about the appropriate baseline for doing contractarian moral and political theory.

But this time—and as importantly—I want to explain

- 2) what otherwise laudable heuristic has led everyone else astray.

And, finally, insofar as any account of how people are led astray always invites

- 3) an account of how the person giving the account wasn't himself led astray.

In other words, what makes me so special? I want to account as well for my own—how shall I put this?—anomalous insightfulness. I begin, then, with ...

## **2. Baselineing**

### *2.1. The Debate So Far*

According to Narveson—and on this score, at least, I have no cause to nay say him—talk of ownership is parasitic on talk of rights. (And/or their correlatives, i.e., duties.) More particularly, to say I own something is just to say I have a right that you not interfere in my making use of it. (Or, correlatively, that you have a duty not to so interfere.)

Rights-talk falls into the same locution set as oughts. To say that I have a right to something is just to say you ought not to interfere with my making use of it. To say that you have a duty to not interfere in my making use of something is just to say you ought not to interfere in my making use of it.

Well, fair enough. But now all the work will be done by our answer to: what counts as interference?

Suppose I grant that you own your car. But while you're shopping I park mine a millimeter from your front bumper, and my friend parks his a millimeter from your back bumper. Clearly, we're preventing you from making use of your car. I should pull ahead, or my friend should pull back, to let you out.

But, you might say, that's because you were there first. My friend and I did something to prevent you from making use of your car.

Fair enough. So now suppose we'd done nothing of the sort. He and I each parked our cars and then yours was craned into place between us by some prankster, a place from which—save by one of us pulling ahead or back—it cannot be extricated. True, it's the prankster who's liable. But suppose he's

---

<sup>13</sup> *Philosophy and Game Theory Meet Each Other*, as yet an unpublished text.

nowhere to be found. So, are my friend and I interfering with your making use of your car or aren't we?

If you say 'yes,' then the distinction between positive and negative rights and duties—a distinction crucial to the libertarian case—has just collapsed. For one species of prankster who, though liable, can't be made to pay, so to speak, is God. And one kind of prank He might play on me is cancer. And so if you don't lend me assistance in combating my cancer you're interfering in my making use of my body.

But if you say 'no'—no, you're not violating my right to make use of my car by refusing to pull ahead or back, and so no, you're not violating my right to make use of my body by not assisting me in combating my cancer—then any motivation one might have had for subscribing to the non-interference interpretation of a right has just evaporated. For what people need from each other is that one pull ahead or back when the other's car is blocked. What people need from each other is assistance in combating one's cancer. Or crop failure. Or what have you. In fact, were it not for these needs, we'd never have been motivated to exit the state of nature/war and enter civil society in the first place.

Not so, says the libertarian. What we need from each other first and foremost is assurance that we won't kill each other. Then and only then are we free to enter into whatever further arrangements might be to our mutual advantage, arrangements like assistance in times of need. That is, assure me first that, regardless of what further arrangements I do or don't agree to, I won't die by your hand, and then we'll talk. Why only then? Because then and only then can those further agreements be truly said to be entered into voluntarily. And why is it important that they be voluntary? Because if and only if they're voluntary does it make sense to say they're mutually advantageous and therefore rational to subscribe to.<sup>14</sup>

I entirely agree. But an agreement not to kill each other is of no use to me if I'll as surely die of my cancer or crop failure. So I propose the following: if you agree to assist me in combating my cancer or crop failure, then and only then will I agree not to try to kill you. This arrangement is clearly to our mutual advantage. And, as clearly, if entered into it would be entered into voluntarily. After all, you're entirely at liberty to reject the deal and take your chances on the battlefield. In fact, even if, having bested you in combat, I have your life in

---

<sup>14</sup> This isn't quite right, since I might have an interest in which I take no interest. For example, unbeknownst to me the car you're offering to sell me really is the deal of the century. Narveson acknowledges such cases as readily as I do. But we're of a mind that the normative political theory based on interests rather than preferences presents epistemic challenges that cannot be met. So, on this privileging of preferences rather than welfare, there's no dispute between us.



my hands, you're still at liberty to choose death over acquiescence. People do so all the time.<sup>15</sup>

Put another way, imagine a world in which life is so "solitary, poore, nasty, brutish, and short"<sup>16</sup> that the difference in utility between life and death is insignificant, but the difference in utility between

- a) life in a state of nature or death and
- b) a compact for mutual assistance

is enormous. Then the case for the lexical priority of a compact not to kill each other would evaporate.

And indeed it's precisely because the difference in utility between life in a state of nature and death is insignificant but the difference in utility between

- a) life in a state of nature or death and
- b) a compact for mutual assistance

is enormous, that Hobbes is anything but a libertarian. For, notes he, a state of nature is a state of war. And in

a time of Warre, where every man is Enemy to every man ... there is no place for Industry, because the fruit thereof is uncertain, and consequently no Culture of the Earth, no Navigation, nor use of the commodities that may be imported by Sea, no commodious Building, no Instruments of moving and removing such things as require much force, no Knowledge of the face of the Earth, no account of Time, no Arts, no Letters, no Society, and which is worst of all, continuall feare, and danger of violent death.<sup>17</sup>

The libertarian thinks we could be motivated to enter civil society even if all we could expect from doing so was escape from this last condition, i.e., danger of violent death. But this is manifestly false. The history of the human race just

---

<sup>15</sup> Elsewhere I've opined (*Philosophy and Game Theory Meet Each Other*, as yet unpublished) that there are only three uses one person can make of another without deferring to that other's cooperation: as the object of perverse sexual gratification, as rather unpalatable live food, and as catapult fodder which, for some reason, one wishes to announce its own trajectory. So the concept of coercion, I claim, is really a con-concept, i.e., a rhetorical flourish designed to con one's interlocutor into accepting as a judicata that which is precisely at issue.

<sup>16</sup> Hobbes, 186.

<sup>17</sup> Hobbes, 186.

is the history of our willingness to risk violent death in exchange for a chance to increase our standard of living.<sup>18</sup>

Most moral and political philosophers will grant all this but insist nonetheless that there's a perfectly serviceable distinction between 'agreements' entered into only at the point of a gun and (let's call them) 'agreements-properly-so-called.' Intuitionists about this distinction are content to leave it at that, i.e., as intuitively obvious. Reductionists, by contrast, are prepared to put in a little effort here.

For example, Narveson—along with Robert Nozick and David Gauthier—reduce the distinction between voluntary and coercive to the distinction between bettering and worsening. There's nothing coercive about my declining to better your situation, say they, so any agreement entered into by which you hope to better your situation is voluntary. By contrast, what we mean by coercion is threatening to worsen your situation.

This is helpful, but not very. For two reasons. First, bettering and worsening require a baseline. And second, even if the bettering/worsening distinction did capture what we mean by the voluntary/coercive distinction, it doesn't tell us what, if anything, is wrong with coercion.

Let's deal with the baselining problem first.

## 2.2. *The Right Baseline*

Some people think—and Narveson, Nozick, and Gauthier seem to be among them—that a condition of our sitting down and bargaining over who gets what share of some cooperative dividend is that we each recognize the other's entitlement to what she brings to the table. This is a mistake. It wouldn't be a mistake if by one's 'entitlement' to something is meant nothing more than that one seems to have control of it. For then all one is saying is that a condition of bargaining is that we each recognize what the other has control over. (Well, yes, since it's hard to imagine my bargaining with you for something that I don't think is within your power to deliver, that's just a 'duh.')

But it's a mistake to suppose we need to recognize each other's ownership of the objects of this control. For to suppose that I need to recognize your ownership of an object is, as we saw at the outset of Section 2.1, to suppose I need to think that I ought not to interfere in your making use of it. But this makes no sense. If I can't interfere in your making use of

---

<sup>18</sup> A similar mistake is made by Rawls, who thinks rational self-interested agents, negotiating from behind a veil of ignorance as to their natural and social endowments, will opt for maximal liberty consistent with a like liberty for all—the so-called Liberty Principle—and a Pareto-optimal distribution of material dividends, subject to the proviso that these positions of inequality be equally accessible to all—the so-called Difference Principle. But, insists Rawls, the Liberty Principle takes lexical priority over the Difference Principle. Rawls is wrong. Rational agents trade liberal for material dividends all the time. And rightly so. So liberal and material dividends are of a piece. Likewise of a piece, then, are life and the commodities of life.

an object, then whether I ought or ought not to doesn't arise. So to say that I ought not to interfere in your making use of an object implies that I *can* interfere in your making use of it. But, if I can interfere in your making use of it, in what sense are you the one bringing it to the table?

Another way to put this is to point out that those who suppose bargaining over ownership requires a prior recognition of ownership are saddled with an infinite regress problem. But no such problem arises if we suppose bargaining requires not recognition of anything, but rather and only the realization of who controls what. It's not that I own my superior marksmanship that will, given plausible bargaining norms, entitle me to a lion's share of the hunt. It's that I control my superior marksmanship. It's that you have to defer to me for its deployment. So the baseline isn't any of the suggestions proffered by Narveson, Nozick, Gauthier, or any other variation on the Lockean Proviso. It's whatever would be the case in the absence of an agreement to the contrary, given what I can bring to the table.

But—and this is the crucial 'but'—what would be the case in the absence of an agreement to the contrary is not that each of us would just return to our respective corners of the forest and carry on hunting on our own. For what I can bring to the table is not just my capacity to better your situation. It's also my capacity to worsen it. So the fact that I'm a better marksman than you means that were we to return to a state of war I'd be more likely to kill you than you would to kill me.

Gauthier has argued that, since a return to a state of war would be sub-optimal, and would be known to be sub-optimal, any threat to return to a state of war would be disingenuous, and so would be regarded as such. So, says he, our military endowments can be discounted. But this is an egregious error in Gauthier's reasoning, made all the more egregious by its inconsistency with his own reliance on the notion of a 'pre-commitment strategy.' For from the fact that it would be sub-optimal, and so irrational, to return to a state of war, it does not follow that it would be irrational to pre-commit oneself to return to a state of war if one doesn't get his way in the bargaining process. So the threat to return to a state of war need not be disingenuous. And so the worsening effects of war are of a piece with the bettering effects of cooperation.<sup>19</sup>

Or put another way, suppose separately you and I can make 4 and 2 widgets respectively but together we can make 10. The cooperative dividend of 4 can be divided to yield take-homes of 7:3, 6:4 or 5:5. But which will it be?

---

<sup>19</sup> Lest I be accused of siding with the strong over the weak, I should point out that pre-commitment strategies can be wielded by the latter as readily as by the former. One could, for example, try the Gandhian ploy. 'Yes,' you might say, 'by killing me I lose 4 widgets. But you lose 1. But it's really only your bottom line that you care about. So I dare you to go ahead and kill me.'

Some people think distribution should be a function of contribution. (This sometimes masquerades as desert.) There is, of course, considerable debate about how to measure contribution, a.k.a. ‘desert.’ Some think that

**measure # 1:** our counterfactual productivities suggest you’ve contributed more than me. So to the baseline of 4 and 2 we should add 3 and 1 respectively, yielding 7:3.

Others think that

**measure # 2:** it’s obviously the cooperation that’s doing all the work here, and so since we’re equally cooperating we should equally benefit from the cooperative dividend. So to the baseline of 4 and 2 we should add 2 to each, yielding 6:4.

Still others think that

**measure # 3:** 4 and 2 serve as a baseline only for whether there’s anything to be gained by cooperating, and so whether it’s mutually rational to cooperate at all. But if it’s rational to cooperate, then the original baseline ceases to be relevant. Since we’ve equally cooperating, we should each equally benefit from the cooperative enterprise, i.e., we should each get 5.

I have no objection to this idea of contribution. Not even if you instead call it ‘desert.’ But I think my interlocutors are being unfairly selective about what counts as such. Suppose, as before, separately you and I can make 4 and 2 widgets respectively but together we can make 10. But by this ‘separately,’ I mean if left to our own devices. If not left to our own devices, you can produce 0 and I can produce 1. How so? Because if not left to your own devices you can’t make any widgets at all. Why? Because you’d be dead. Whereas if I’m not left to my own devices I won’t be dead—because, suppose, I’m a better warrior than you—though now I can make only 1 widget, having expended the other in killing you. So why isn’t 0 and 1, rather than 4 and 2, the right baseline from which to begin negotiating?

In fact, I claim that 0 and 1 is the right baseline from which to begin negotiating, because 0 and 1 is what would be the case in the absence of an agreement to the contrary, given what we can each bring to the table.

### *2.3. The Ethics of Worsening*

It would strawman my position to reduce it to ‘To each according to his threat potential!’ Just as nations can adopt a scorched earth policy to slow the enemy’s advance, likewise can they adopt such a policy to demotivate the invasion in the first place. But that means they’re also bringing their productive capacity to the table. And the same is true of interpersonal negotiations. No one—save perhaps Rawls, insofar as productive capacity is a natural endowment—thinks that productive capacity should be discounted in interpersonal negotiations.

Why, then—apart from the disingenuousness argument already refuted—would anyone think military capacity should be discounted?

One could think this, I submit, only by importing a moral intuition, i.e., the intuition that it's wrong to profit by threat of violence. But if we allow a moral intuition to inform our theory we've forfeited our claim to full reduction. If there is such a moral intuition—and I don't deny there is—then it's to be explained and, if appropriate, justified, by non-moral premises. In other words, we have to demonstrate how the discounting of my capacity to worsen your situation can arise from the counting of my capacity to worsen your situation.

I turn to that demonstration now. As it turns out, however, in doing so I'll be hitting two birds with one stone. For my account of how the discounting of my capacity to worsen your situation arises from the counting of my capacity to worsen your situation will itself arise from my account of ...

### 3. Why Everyone Else is Wrong about Baselineing

#### 3.1. *The Shallow Analysis*

There are a number of ways to characterize the mistake that Narveson, Nozick, and Gauthier are making in supposing the baseline for bargaining is some species of the Lockean Proviso. It could be characterized, as we've already seen, as an illegitimate importation of a moral intuition. It could be characterized, as we've already seen, as a failure to appreciate the rationality of irrationality, i.e., of employing a pre-commitment strategy. But I think the most instructive way to characterize it is as a failure to press one's own reductionist program to completion.

To be fair, in this timidity, Narveson, Nozick, and Gauthier are in good company. For Hobbes himself makes the same mistake. In Chapter XIII of *Leviathan* Hobbes takes pains to make clear that

To this warre of every man against every man, this also is consequent; that nothing can be Unjust. The notions of Right and Wrong, Justice and Injustice have there no place. Where there is no common Power, there is no Law: where no Law, no Injustice. Force and Fraud, are in warre the two Cardinall vertues ... It is consequent also to the same condition, that there be no Propriety, no Dominion, no Mine and Thine distinct; but onely that to be every mans that he can get; and for so long, as he can keep it.<sup>20</sup>

And yet in Chapter XX Hobbes declares that

In a condition of meer Nature ... the right of Dominion over the Child dependeth on [the] will [of the Mother], and is consequently hers. [For] seeing the Infant is first in the power of the Mother, so as she may either nourish, or expose it, if she nourish it,

---

<sup>20</sup> Hobbes, 188.

it oweth its life to the Mother; and is therefore obliged to obey her, rather than any other; and by consequence the Dominion over it is hers.<sup>21</sup>

Hobbes' use of "oweth" and "obliged" could be regarded as a harmless slip of the tongue if, as with my comments about 'entitlement' at the outset of Section 2.2, all he means is that a child realizes what resources its mother controls. But, if that's all Hobbes means by "oweth" and "obliged," then all he'd be saying is that, since

- 1) the child realizes what resources its mother controls,
- 2) the child realizes what resources its mother controls.

So, for his claim to be synthetic rather than merely tautological, he must mean something else by "oweth" and "obliged."

And indeed he does. He means that the child is now under a moral obligation to obey its mother, i.e., to obey her even after it ceases to be dependent on her. He means that the child owes her this obedience in repayment for the benefit she has bestowed upon it. And yet this is impossible on Hobbes' own account. For in Chapter XIV he makes clear that

If a Covenant be made, wherein neither of the parties performe presently, but trust one another; in a condition of meer Nature ... upon any reasonable suspition, it is Voyd ... For he that performeth first, has no assurance the other will performe after; because the bonds of words are too weak to bridge mens ambition, avarice, anger, and other Passions, without the feare of some coerceive Power ... And therefore he which performeth first, does but betray himselfe to his enemy; contrary to the Right (he can never abandon) of defending his life, and means of living.<sup>22</sup>

Moreover, this contradiction between Chapter XX and Chapters XIII and XIV is repeated within Chapter XIV. For there he says that

Covenants entered into by fear, in the condition of meer Nature, are obligatory. For example, if I Covenant to pay a ransome, or service for my life, to an enemy; I am bound by it. For it is a Contract, wherein one receiveth the benefit of life; the other is to receive mony, or service for it; and consequently, where no other Law ... forbiddeth the performance, the Covenant is valid. Therefore Prisoners of warre, if trusted with the payment of their Ransome, are obliged to pay it.<sup>23</sup>

Here there can be no suspicion that by being bound by such a covenant all Hobbes means is being well-advised to comply with it. For under the circumstances

---

<sup>21</sup> Hobbes, 254.

<sup>22</sup> Hobbes, 195.

<sup>23</sup> Hobbes, 198.

described one would be ill-advised to pay such a ransom. In fact, one would be downright crazy to pay it! And yet Hobbes thinks one should pay it. He should pay it because, as with the child example, he received a benefit. But this contradicts Hobbes' earlier denial that there can be moral obligations in a state of nature. Indeed, that moral utterances can have any meaning in a state of nature.

How could Hobbes have made such an obvious mistake? There are only two possibilities. The first is that he's not making a mistake. We've been making a mistake in supposing Hobbes is, well, a Hobbesian, whereas, in fact, he's a natural law theorist. On this reading, Hobbes thinks that one is obligated to "performe [one's] Covenants made," even in a state of nature, but that on, and only on, "reasonable suspicion" that the other will not perform his side of the bargain, one is relieved of that obligation. And that this "reasonable suspicion" is satisfied by there being "no assurance the other will performe after."<sup>24</sup> But in both the ransom and child cases that assurance is given by the other having performed his side of the bargain already. But, if Hobbes thinks one can have obligations in a state of nature, he must be a natural law theorist, and not, as we've always thought, the father of moral and legal positivism.

But hold on. On this view of Hobbes' view, the institution of sovereignty is required not to create law but to enforce it. But that's just Locke's view. And so the contrast between Hobbes and Locke—a contrast that virtually defines political philosophy—collapses. But, in that case, rather than allow the contrast virtually definitive of political philosophy to collapse, wouldn't it be better to simply consign Hobbes to the flames on ground of redundancy and then make up a new person named Hobbes\* to be the torch-bearer for moral and legal positivism?

The second possibility is that Hobbes is making a mistake—or, if you prefer, that Hobbes\* is making a mistake—both in the child case and in the ransom case. It's the mistake anyone makes when involved in a paradigm shift. From time to time, and until the shift is complete, one is prone to slip back into the old way of thinking.

If this is right—and I think it is—then the mistake my interlocutors are making is they're having trouble sustaining the shift in our paradigm from natural law to positivism. This is not to suggest natural law theory is the natural way to think and positivism is the stretch. On the contrary, natural law theory is a highly unnatural way to think about law. But it's an unnatural way to think about law that's taken hold of us, and will not easily loosen its grip.

But why has it taken hold of us? Why won't it easily loosen its grip? Answering these questions requires ...

---

<sup>24</sup> Hobbes, 195.

### 3.2. *A Somewhat Deeper Analysis*

My hypothesis is that we—by which I mean everyone except me—have been mistaking a mere heuristic for the real thing. This need not surprise us. We do this all the time. (And this time by ‘we’ I do include myself.) For example:

Some crimes, though still crimes, aren’t harmful, e.g., smoking marijuana. But most are. So, when we talk about the crime rate going up or down, we take ourselves to be talking about harm increasing or decreasing. We could cut the crime rate in half tomorrow by decriminalizing marijuana. But no one would think we’d have thereby halved the harm in our society. So it’s a mistake to equate crime with harm. But it’s not a mistake to have adopted this equation. For only occasionally will this equation lead us astray.

Similarly, then, the vast majority of our decisions are made under conditions of civil society. So vast, in fact, that for some fortunate people the state of war has become so remote that they’ve ceased to be able to imagine it, and so they’ve ceased to be able to think about what dispositions would no longer be appropriate under such circumstances. Thus, the dispositions inculcated in them in and by a state of civil society go with them when they enter—what, for them, can only be the thought experiment of—a state of nature. And so in the thought experiment they make mistakes. Hobbes mistakenly thought that the child owes its mother something and that the prisoner should pay his ransom. Narveson and Gauthier mistakenly think we should discount our military endowments. And so on.

These are relatively harmless mistakes. They eventuate in poor advice, but because this advice violates (what Steven P. Lee calls) the Principle of Tolerable Divergence, no one takes it seriously anyhow. What’s not so harmless, however, is actually being unable to toggle from dispositions appropriate to civil society to those appropriate to a state of war. In fact, it was arguably this inability which cost six million Jews their lives.

But if forfeiting one’s ability to toggle from dispositions appropriate to civil society to those appropriate to a state of war is liable to cost one his life, anyone who forfeits that ability must be, well, just stupid. But, as we saw in Section 1.4, that the Jews of Europe were just stupid is a bad error theory. Clearly, then, we need ...

### 3.3. *A Deeper Analysis Still*

So far I’ve conjectured that the predilection for discounting one’s ability to worsen the other’s situation in fixing the baseline for negotiations is the product of an otherwise laudable heuristic assumption that we’re already in a state of civil society, and so our capacity to worsen the other’s situation is already off the table. But, insofar as the prospect of returning to a state of war is never so far removed that we can afford to discount it, there must be some value to this discounting—beyond its savings in computational resources—that compensates for the vulnerability it generates. What might that be?



I should like to propose that this discounting is a species of self-effacement.<sup>25</sup> To explain: a mental state—be it a hope, a fear, a belief, a set of beliefs constituting a theory, or what have you—is self-effacing just in case one consequence of entertaining that mental state is that one is well-advised to cease to entertain it, if she can—call this ‘weak self-effacement’—or to entertain its opposite—call this ‘strong.’ For example, one’s desperation to ‘meet’ someone of the opposite sex can often be weakly self-effacing. So, could it be that touting one’s capacity to worsen the other’s situation is, in significantly many cases, weakly—or perhaps even strongly—self-effacing? That, in other words, there are significantly many circumstances under which we fare better in our negotiations by transparently forfeiting our conditional intention to return to a state of war if things don’t go entirely our own way. (Note the importance of the transparency condition here.)

Such situations are, I contend, legion. Most cooperative enterprises—whether hunting, making widgets, or just sharing a bed for warmth—involve placing ourselves in a position of physical vulnerability to our cooperative partner. An arrow meant for the deer can as readily pierce me. A hammer for widget-making can as readily be used as a weapon. The clock radio readily at hand by which I can dispatch you is as readily at hand to you to dispatch me. But we cannot simultaneously attend both to the enterprise and to our own protection. In other words, to “solitary, poore, nasty, brutish, and short,” Hobbes could have added ‘sleepless.’ So we have to be able to mutually self-efface our conditional intention to return to a state of war—or more generally our willingness to worsen the situation of the other rather than just decline to better it—since otherwise cooperation becomes impossible.

To self-efface a mental state is to render it virtually unrecoverable. For if it can be too readily resurrected—if its self-effacement is only surface—that it’s only surface will be transparent to one’s co-player, and so she cannot let down her guard, and so she cannot cooperate. But the same is true of the first player. She can’t cooperate unless she has transparently self-effaced her conditional intention to resort to violence should she not get her way. (Hence the importance of the mutual transparency condition.)

But, one might argue, since the intention to resort to violence (a.k.a. ‘worsening’) is only conditional, so long as we each ensure that the antecedent is never satisfied we’re assured that the consequent won’t be either. So why the need for self-effacement? Because one can’t negotiate—at least not at close quarters—without assurance that one won’t be killed (or more generally one’s situation worsened) in the course of the negotiations. So even to place ourselves in sufficient proximity to each other to enter into negotiations, we need to mutually and transparently self-efface our conditional

---

<sup>25</sup> To the best of my knowledge, this notion of ‘self-effacement’ was first introduced into philosophy by Derek Parfit in *Reasons and Persons*.

intention to return to war/worsening should we not come to some kind of accommodation.

Hence Gauthier's insistence that in bargaining we have already forgone resort to violence, and so threat advantage can be discounted, and Narveson's insistence that the forgoing of violence takes lexical priority. In other words, they're not so much wrong as they are mistaken in why they're right. They're right because they're reporting on, and reflecting, the bargaining norms of people who've self-effaced the conditional intention to resort to violence/worsening. So ...

### 3.4. *What Makes Me So Special?*

Just that, having written my doctoral dissertation, lo these many years ago, on the logic of self-effacement, I'm now hardwired to look to self-effacement to explain phenomena that would otherwise be paradoxical.

I'm also special—or so I'm told—in that I'm at least slightly more psychopathic than most people, and so the pre-effaced state enjoys some residual resonance in me, whereas in my less psychopathic friends and colleagues the self-effacement is more complete. I am, in short, Hobbes' Foole. But it's my very moral Foolishness that gives me the philosophical leg up on my moral betters.

## References

- Gauthier, David.  
 1986 *Morals by Agreement*. Clarendon.
- Held, Virginia.  
 "Non-Contractual Society." *Canadian Journal of Philosophy, Supplementary Volume 13*: 111 (1987).
- Hobbes, Thomas.  
 1968 *Leviathan*. Edited by C.B. Macpherson. Pelican Books.
- Hume, David.  
 1998 *An Essay Concerning the Principals of Morals*. Edited by Tom L. Beauchamp. Open Court Pub. Co.
- Hume, David.  
 1975 *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge, 2<sup>nd</sup> ed. Revised by P.H. Nidditch. Clarendon.
- Locke, John.  
 1980 *Second Treatise of Government*. Edited by C.B. Macpherson. Hackett Publishing Company.
- Lee, Steven P.  
 1993 *Morality, Prudence, and Nuclear Weapons*. Cambridge University Press.
- Mill, J.S.  
 2007 *Utilitarianism*. Dover Publications, Inc.
- Narveson, Jan.  
 2001 *The Libertarian Idea*. Broadview Press.

Nozick, Robert.

1974 *Anarchy, State, and Utopia*. Basic Books.

Parfit, Derek.

1984 *Reasons and Persons*. Clarendon.

Rawls, John.

1971 *A Theory of Justice*. Belknap Press of Harvard University Press.

Viminitz, Paul.

*Philosophy and Game Theory Meet Each Other*. Unpublished.