IDEAL THEORY

David Schmidtz

There is a particular kind of ideal theory that does not work: namely, theorizing about ideally just worlds. Such theorizing fails to bear on what we have reason to regard as an ideal response to the practical concerns to which justice is a response, and instead bears only on what *would* be ideal if we were to depart far enough from the human condition as it really is.

This chapter is a critical assessment of, and a gesture at an alternative to, recent work on ideal theory. The following chapter continues the task of characterizing an alternative kind of ideal theory, *realistic idealism*, in which an ideal is understood to be, at very least, a disciplined response to a real problem.

CHAPTER I. IDEAL THEORY: WHAT IT WAS

Here is the idea that launched a debate and defined ideal theory.

> The intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life. My main concern is with this part of the theory. Nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen; only then do the parties ask which principles to adopt under less happy conditions. (Rawls 1999: 216)

Rawls goes on to say,

> With the presumption of strict compliance, we arrive at a certain ideal conception. When we ask whether and under what circumstances unjust arrangements are to be tolerated, we are faced with a different sort of question. We must ascertain how the ideal conception of justice applies, if indeed it applies at all, to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice. The discussion of these problems belongs to the partial compliance part of nonideal theory. (1999: 309)

As Rawls conceded, "Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life" (1999: 8). Yet, Rawls adds, "I consider primarily what I call strict compliance as opposed to partial compliance theory" (1999: 8). Why? "The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. . . . At least, I shall assume that a deeper understanding can be gained in no other way" (1999: 8).

As Rawls knew, ideal and nonideal theory are not rivals. Each plays a role in political

theory done well. Yet, as Rawls also knew, where we start is critical. Assuming perfect compliance leads down one road. Realistic assumptions lead down another. We assume perfect compliance not because more realistic assumptions go nowhere so much as because they do not go where Rawls wants theorizing about justice to go. As Rawls puts it, "We want to define the original position so that we get the desired solution" (1999: 122). We can see why he would say that. However, precisely because Rawls is right that it matters where we start, we cannot afford to be uncritical about where he starts.

Rawls also says, "until the ideal is identified, at least in outline—and that is all we should expect—nonideal theory lacks an objective" (1999a: 90). Really? Alexander Graham Bell had a sense of where long-distance communication was in his day and where it could go. Thomas Edison had a sense of the possibility of recorded sound. They did not lack for objectives, and their objectives were rooted in their sense of what was possible, not in any sense of what was ideal. They did not need a theory about whether ideal sound transmission would be faster than the speed of light. Progress in solving problems was relevant and meaningful; perfection was neither.

Where there are facts, subject to change in ways that matter, and where there is something we can do, we have a problem. Where there are problems, there can be objectives. Where there are objectives, there can be theorizing with a point. Problems set parameters for what to count as a solution, and ultimately for what would be our best—our ideal—response.

Likewise, some of our best legal reasoning comes from judges aiming, not to approximate perfection, but to solve a problem—to resolve disputes brought before their bench by real litigants. It is not law's purpose to be an ideal answer to all possible questions. Systems evolve, new conflicts emerge, and judges will one day need to say more than has yet been said about what would make a newly emerging kind of dispute easier to resolve or avoid in the future. Real solutions do not track perfection but instead play a perpetual game of catch-up with evolving problems.

Realistic idealism is theory about ideals, but does not start with ideals. Realistic idealists know that it matters where we start, so they do not *start* with a desired conclusion. Instead of letting an *answer* dictate what they count as a question, theorists who succeed at being realistic let real, located, practical questions dictate what they count as an answer.

Realistic idealism does not give up on ideals; it gives up on *starting* with ideals. (See

Galston 2010 and Elkin 2006.) To those who start with ideals, it can seem okay to have an intuition that everyone ideally would have the same income. Perhaps it is only human to wish such intuitions were untestable—fact insensitive—in a world that has indeed tested those intuitions and found them wanting (to the extent that worlds of uncontrolled variables ever test the hypotheses of social science). But realists who seek ideal responses cannot look at it that way. To them, there is a general rule: for any problem P, there may be an ideal solution, but for S to be an ideal solution to problem P, S must first be a solution to problem P. From the perspective of a realistic idealist, it is not ideal to set aside the problem, and it is not ideal to set aside whether S solves the problem.

Suppose we call something ideal. Then we pause to wonder whether there is any reason—any reason whatever—to call it ideal. Suppose Thomas Edison says, "in an ideal world, our incomes would be equal, and there would be no carnivorous plants." Under what circumstances would that answer speak to a real question? Here is a suggestion: theorizing goes astray, not when it bears on ideals, but when it *fails* to bear on problems. When it fails to bear on problems here and now, theorizing fails to bear on what we have reason to regard as an ideal response to here and now, and instead bears only on what *would* be ideal if we depart far enough from the human condition as we know it. Realistic idealism is theorizing about ideal *responses*. Realistic idealism identifies *x* as worthy of aspiration here and now, starting from a sober assessment of real problems here and now. Disciplined idealism—realistic idealism—manifestly is worth doing well.

II. WHAT IT COULD BE

In an ideal world, would there be progress, or would progress be inconceivable? It depends on whether an ideal society is an ideal outcome or an ideal process. Let's say a *utopian* ideal is a "peak" with respect to which further progress is inconceivable; by contrast, a *realistic* ideal is a process that makes progress likely. A utopian, seeing problems, imagines how nice it would be not to have them. A realist, seeing problems, imagines what would solve them.

Suppose we define a perfectly just world as one that could not be more just. What else would a perfectly just world be? Utopian idealism, imagining a world so perfect that nothing remains for justice to demand, is one alternative. Realistic idealism, by contrast, consists of

imagining a perfect response to circumstances that demand a response. Let me stress: utopian theory and ideal theory are not the same thing. To suppose that only utopian theory counts as ideal theory is to pay utopian theory a complement that it does not deserve.

To elaborate further, the best available response is not always what we would call ideal. We call the best available response ideal only when we accept some fairly strong version of the thought that we could not have done better. Suppose I am imagining that cooking lasagna would be ideal for our dinner tonight. Then you mention that you are allergic to tomatoes. I should reply, "That fact calls for a change of plan, because given that fact, serving lasagna tonight would be far from ideal."[1] Suppose instead that I reply by saying, "So far, there is no discernible defect in my plan to make lasagna for our dinner tonight. It is you that is defective, not my lasagna." Why does this sound childish? Perhaps because it disregards the fact that there are people for whom my lasagna would not be ideal, and you are one of them. The problem is not that lasagna is infeasible but that it is undesirable under the circumstances. If I find that I literally cannot execute my plan (perhaps tomatoes are unavailable), then I am finding that serving lasagna is not feasible. By contrast, if I find out what a bad idea it would be to serve you lasagna, then I am finding that serving lasagna is undesirable. Being infeasible does not entail that I should stop calling my plan ideal. But being undesirable does.

Because utopian idealists aim to describe an ideal world rather than an ideal response to the world, it follows that what they call *ideal* does not entail *feasible.* So, far, there is nothing to quarrel about. However, note how much more peculiar it is if what they call ideal does not even entail *desirable*. And yet, whether S is desirable is as fact-sensitive as whether S is feasible.[1]

Some theorists concede that we cannot, on humanitarian grounds, put (their intuition about) justice into practice.[2] But if a vision is of a system *characteristically* at odds with humanitarianism, then the problem with the envisioned system is not that we lack any way to *get* it, but that we lack any reason to *want* it. If a system's logic makes its inhumanity predictable when premised on human nature as it is, the problem with instantiating that logic is not that we can't do it, but that we shouldn't want to.

Edward Hall, similarly, concludes that

the insistence that political philosophy must start with the acceptance of various facts, such that in politics people do not display the sort of unity of purpose that they do in Cohen's camping trip, is not a 'feasibility' concern at all, but rather the more fundamental requirement that we actually address the

---

[1] NOT SELF-CONTAINED

practice with which we claim to be concerned. (2013: 180)[3]

But again, this is not a broadside against ideal theory. My target here is a specific kind of ideal theory: a utopian idealism that is not even trying to discover an ideal response to a recognizably human problem.

Much of what we currently call ideal theory is an exercise in imagining how we would reinvent the world if only we could start with a clean slate and did not need to start from here. In truth, we live in a Neurath's boat and cannot do better than to replace one plank at a time. What actually needs doing had better be guided by ideals of a kind, but if we think we need to be guided by a vision of the ideal boat, we are mistaken as a matter of plain fact.

Real solutions to real problems characteristically exhibit twists of historical contingency that armchair philosophy cannot anticipate. The convention of driving on the right solves a problem in some societies. It makes no difference whether driving on the right is a necessary condition for solving a problem; it makes no difference whether a convention of driving on the left would have been at least as good. When a convention of driving on the right emerges and solves an important problem, members of that society become warranted in expecting each other to drive on the right.

CHAPTER III. TOWARD A PHILOSOPHY THAT STARTS AT THE START

In practice, to begin with a problem is to begin more problematically with a *description* of a problem. Descriptions are idealizations. All descriptions set aside details, notwithstanding the fact that what gets set aside as a detail is not guaranteed to turn out to be merely a detail.

A theory articulates an understanding, which entails deciding what to ignore. Although we face the human condition in its full complexity, theorizing begins with describing, and, as Peter Godfrey-Smith observes of biological theories, "ignoring some features in a description of a system is inevitable to some extent in any description" (2009: 48). As Alan Hamlin observes in his chapter, a map of the London Underground sets aside nearly everything about London, even distances and scale, so as to distill the one kind of information that the map's users seek to glean from it, namely the sequences of stops making up the network's lines.

Yet, however controversial our description inevitably must be, there remains such a thing as the human condition. Simon Hope (2010: 135) warns against working on idealized

representations with cherry-picked features rather than on the situated problems of real human beings. But as Hope also notes, the bare fact of articulating a question is not enough to convict us of begging the question. There is no *uniquely* accurate account of the human condition, yet there are more and less accurate accounts.[4]

When Rawls agrees with Hume that we live in a world of real but limited generosity, we see the truth in what they say. They are not telling us everything of any conceivable relevance—of course they are simplifying—but neither are they making it up. Second, Hume and Rawls agree, we live in a world of scarcity. Third, the scarcity we face is manageable if we cooperate. Fourth, we are intelligent; cooperation is possible. None of us can know everything, decide everything, or do everything. Yet, we anticipate and respond to each other, communicate, give our word, by our deeds make our word count for something, and in time learn to trust one another. These are enduring features of the human condition. They motivate us, and enable us, to be a kingdom of ends.

Rawls says society is a cooperative venture for mutual advantage. Hume saw that, as Rawls would concede, the emergence of a cooperative venture is an achievement. Thus, Hume asked, which basic structures *foster* society as a cooperative venture? Hume saw this as an empirical question and saw social science as a key to real answers. He sought "to introduce the Experimental Method of Reasoning into Moral Subjects."[5]

Of course, neither Hume nor Rawls represented their maps of the human condition as complete. Maps at best supplement rather than replace good judgment. Maps *represent* a problem. They do not *constitute* it. Moreover, maps can mislead. Thus, a challenge for any theorist is to distinguish what can safely be set aside as a distracting detail from what cannot.

Rawls launched his greatest work with the thought that justice is the first virtue of social institutions (1999: 3), suggesting that a theorist's main task is to articulate principles of justice. Rawls's sentence resonates. It is lyrical, poetic, compelling.

But it is not right. To Bernard Williams, the first institutional question concerns not justice so much as "the securing of order, protection, safety, trust, and the conditions of cooperation. It is 'first' because solving it is the condition of solving, indeed posing, any others" (Williams 2005: 3). Historically, we make progress when we acknowledge that the first virtue of social institutions is that they enable us to be neighbors. Being neighbors is not everything, but it is a start. In practice, we first secure a framework of mutual expectation that keeps the

peace well enough to enable society to be, in the most rudimentary and non-theory-laden sense, a cooperative venture for mutual advantage. Then we start talking about justice.[6]

The further implication is that when we get to a point of being able to talk about justice we realize that we have to go beyond Williams' insight that what Rawls calls justice is not the first institutional question[2]. The fact is, what Rawls calls justice is not even the first question of *justice*. That is, Rawls's opening question is about distributing the fruits of cooperation, but before we begin to distribute what separate persons bring to the table, fairness mandates figuring out how to respect the fact that they brought it.[7]

Rawls's thought experiment, like most thought experiments meant to elicit intuitions about distributive justice, has us facing the above circumstances as adults, meeting to divide a pie from nowhere, devoid of historical baggage. In reality, we arrive at different times to a world knee-deep in baggage.[8] We each arrive on the scene to find a pie already produced, divided, claimed, and in use by someone else.

Philosophers for whom the pie's history is nothing but a distracting detail are primed to see equal shares as presumptively fair shares. But except when we actually do arrive simultaneously, there is no fact of the matter regarding what to count as an equal share. Philosophers are trained to set this detail aside, but it changes everything. Suppose 20-year old Smith is hired today at half the wage of his 40-year old colleague Jones but twice the inflation-adjusted wage that Jones was paid when Jones was hired twenty years ago. Does one of them have a valid complaint? If so, which one? Must we avoid progress, lest people born today have a higher quality of life than (and thus an arbitrary advantage with respect to) people born a generation ago? Further, must equal shares *look* like equal shares at any particular moment? Or is it okay for shares to even out over a lifetime, where highly skilled 50-year-olds have incomes dwarfing what unskilled teenagers are making *at that moment*?[9]

Another sometimes pivotal feature of the human condition is that we not only arrive at different times: we arrive helpless, incapable of the reciprocity that grounds relations among free and equal adults. *We arrive to a community that does not yet need us*, and whether it some day will be better off with us than without us is not a given. That we make it to the table at all implies that we had help.[10] Just as there tends to be no fact about what counts as an equal share, there is not always a settled truth about what counts as returning a favor, or to whom a favor is

---

[2] NOT SELF-CONTAINED

or could be owed. Do we have debts to the next generation because of what the previous one gave us?[11] Answers come not from philosophical analysis so much as from contingencies of ongoing mutual adjustment of expectations within communities. I distil this lesson from observed histories of people muddling through, avoiding and resolving conflict. It is not something I would have seen simply by surveying my intuitions. What people around us call justice is not something we lightly set aside as a distracting detail—not even if we fail to see the point of seeing things their way.

Obviously, we cannot settle disputes by deciding that our vision has a right to be colonial and that we can condescendingly dismiss rival visions as unreasonable. Real resolution starts by aiming for real resolution. To be in the grip of a vision—*any* vision—is problematic. What we need is not to envision but to listen. That is, we need politics.

CHAPTER IV. SEPARATENESS (relate to Solipsism)

Treating unconditional giving as an ideal is a way of setting aside rather than honestly addressing the problem of politics. Here is the argument.

The Prisoner's Dilemma models a key insight into the human condition. As described in the chapter on solipsism, declining to cooperate is a dominant strategy. But when we contrive to play repeated games, that enables you to play "tit-for-tat" (moving as your partner moved in the previous round), and in that way reciprocate in the simplest way possible, and thereby make it pay for your partner to cooperate. Although withholding pays more in a one-shot game, reciprocated cooperation pays more than reciprocated withholding in the long run.[12] This bit of game theory can help us to distinguish real from spurious ideals. Arguably, there is an ideal strategy in a repeated Prisoner's Dilemma: namely, reciprocity. Why? Because (1) it matters whether your partner cooperates, yet (2) you do not choose whether your partner cooperates. However, (3) you can make it pay for your partner to cooperate. This logic makes reciprocity an ideal strategy in a repeated Prisoner's Dilemma.

CHAPTER (TRAFFIC MANAGEMENT)

In this context, reciprocity is analogous to defensive driving. It is "driving" to a high

standard: it acknowledges fellow players as players. There is no better strategy for making cooperation pay.[13] By contrast, a spurious ideal of unconditional giving makes free riding pay. It is a recipe for conflict.

It is a mistake to think that solutions to idealized problems can set the bar regarding what to count as having high standards.[3] In our world, driving apt for a world of perfect drivers would be driving, not to a higher standard, but to a grossly inferior one. A genuine ideal portrays driving well in a world of other drivers, which is to say, being a defensive driver. Whether drivers *ideally would not need to drive defensively* has nothing to do with what counts as being an ideal or even a minimally competent driver in the real world. The same could be said of the idea that people ideally would not need to anticipate that the people around them are capable of strategic behavior. Such a conjecture has no bearing on how to respond ideally or even competently to a world like ours.[4]

An ideal traffic manager's idealism rises to the level of seeking ideal responses to real problems. If a proposed solution would not help, the serious conversation is over. There is no further question about whether we can imagine it helping under ideal circumstances.

Setting aside compliance problems goes astray not because it bears on ideals but because it fails to bear on problems. To say "ideally we would not have compliance problems" is like saying "ideally we would not need to drive defensively." It is a remark about a world whose problems are not like ours.

One possible view here is that an ideal of unconditional giving "may play an inspirational role"[14] but "it is not feasible. It is therefore naïve, and ineffective, to hold existing societies to account on the basis of such demanding moral standards."[15] Such a view treats unconditional giving as too much to ask of all-too-human beings. It reaches too high. My view, by contrast, is that treating unconditional giving as ideal asks too little, not too much.

My claim is not that realistic idealism asks *less* than utopian ideal theory asks: on the contrary, it asks more. Realistic idealism is more demanding insofar as it demands that (1) we take responsibility for motivating the people around us; (2) we do not take their motivation for granted, and (3) we do not call them defective just because we have a theory that cannot work without treating them as something that they are not.

---

[3] NOT SELF-CONTAINED

[4] Andrew Mason observes that we are not always sensitive enough to whether theorizing has "identified a principle of justice for a different kind of being rather than a more ultimate principle of justice" (2012: 539).

CHAPTER HIGH STANDARDS (relate to Solipsism)

David Estlund takes as given that idealistic theories posit standards that can be unrealistically high. Surely he is right about that, but he does not consider the corollary that they also can be unrealistically low. He says,

> A hopeless theory can be dangerous, of course. The soundness of the standards might lead some to take actions in their pursuit, and this might be bad. Actions in pursuit of what will never be achieved can be wasteful or even disastrous. A theory that counsels action in pursuit of high standards that are not sufficiently likely to be achieved, where the costs of failing are very high, often deserves to be chastised as utopian. On the other hand, some people might be led by unrealistically high standards to improve themselves or their institutions, even though not all the way (full achievement is hopeless after all, by hypothesis). (2014: 7–8)

My claim: what makes utopian standards every bit as dangerous as Estlund says is how low they are, not how high they are. Revolutionaries treat utopian standards as licenses to ignore what humanity has learned from experience. Thus, to conceive of justice as demanding unconditional giving is to conceive of justice, not as asking too much, but as asking nothing. To come to the table as a reciprocator is to demand that partners cooperate; to instead come as an unconditional cooperator, demanding nothing, invites free riding.[16] That an action guide would be ideal *if* competently implemented by all people together has nothing to do with whether it is ideal in a world where the antecedent does not hold.[17]

Theorists sometimes assume they have high standards, and console themselves with the thought that human nature is too imperfect to live up to their high standards. In truth, the problem is not that other people cannot live up to so-called "high" standards. The simple reality: there typically is no reason why they should. People have visions of their own. Liberalism is the insight that other people having hopes and dreams of their own is not a problem.

CHAPTER V. TOWARD A REAL IDEAL: FOSTERING COOPERATION

Idealization does not relieve us of responsibility for actually defending the substance of what we call justice. G. A. Cohen offers what seems to present itself as a heroic defense of the thesis that not all principles of justice are fact-sensitive. But that is a red herring. No one doubts that there are fact-insensitive fragments of justice. To give an obvious example: punishing the innocent per se is not required by justice. Nevertheless, the fact remains that overall conceptions

of justice are not actually defended until they are defended with reference to facts.

G. A. Cohen supposes that principles are more fundamental than rules of regulation, whereas in reality, the trail of justification leads both ways. We can defend a proposition by arguing that it rests on something deeper and foundational. But to defend a foundation qua foundation is to judge it by what people build on it. When we look at what gets built on a foundation, we enter a world of fact-based empirical contingency. Looking at what gets built on a foundation does not *prove* anything, if by proof we mean being able to *deduce* that the principle necessarily leads to that result. Still, at some point, as a matter of science and common sense, we look at repeating patterns and conclude that this is a mistake we cannot justify repeating. We cannot make a particular conception just, merely by *calling* it just. What Cohen is *calling* justice is an intuition: ungrounded, untested, barely articulated, and massively fact-influenced (by his personal history). Presumably Cohen's conception of justice is internally consistent, as are infinitely many conceptions that experience counsels us to reject.[18] What sorts out conceptions is information—facts—about whether they bear any relation to anything that any situated human being has reason to treat as worthy of aspiration.

To demand an ideal response is, first of all, to demand a response. The problem with unconditional giving is not that it asks too much, but that it is unresponsive to the human condition. Humans are social. They are separate. And for better or worse, they respond. If an ideal response is cooperative, then an ideal way of conducting oneself *among people who decide for themselves* is a way most apt to induce that cooperative response. Unconditional cooperation is not an ideal but a strategy for fostering an actual ideal: namely, cooperation. But whether unconditional cooperation truly fosters the ideal of cooperation under realistic conditions is a matter of observation. Theorizing can help us predict what we will observe, but theorizing does not settle the matter.

We might say reciprocity is realistic while unconditional cooperation is utopian as a strategy for realizing the ideal of cooperation. Rousseau opens his *Social Contract* by announcing that, "we study men as they are and laws as they might be."[19] The part about laws signals that the theory aims to be idealistic; the part about men signals that the theory's idealism aims to be realistic, not utopian. Crucially, an idealism that aspires to identify institutions ideal for men as they are, taking into account how men predictably adapt to favorable conditions, remains for all its idealism a work of realistic political economy. It yields testable predictions

about how institutions work, thus yielding a basis for deciding whether an institution is worth wanting. A theory that ignores facts about human potential can fail to be *idealistic* about human potential precisely because it fails to be *realistic* about human potential.[20]

A work of political apology might, by contrast, defend institutions that observably turn people as they are into monsters, where defense consists of imagining such institutions turning people into angels instead. A political apology imagines utopian conditions under which realistically *indefensible* institutions *would be* defensible. Those institutions are an ideal response to nothing in human experience, but utopians find them inspiring all the same.

When Rawls says society is a cooperative venture for mutual advantage, this is not a description, but a vision of what a society of reciprocators would be like. A real society is more, and also sometimes less, than a mutually advantageous cooperative venture. So, Rawls is telling us about "laws as they might be," not society as it is. Yet, intuitively, there is nothing silly about Rawls's idea—despite a certain lack of realism—that society is a mutually advantageous cooperative venture. Mutually advantageous cooperation is not mere wishful thinking. It is a useful approximation, an accurate (even if simplified) description of society at its best. It is an ideal. But what makes the Rawlsian ideal worthy is not its portrayal of an ideal society, but its portrayal of a society's ideal response to the actual human condition. He is describing what our society looks like when we respond to the actual human condition with the best strategy we have for responding to that condition: namely, with reciprocity.

In our world, reciprocity is exactly right as a response to a defining problem of political theory, namely specifying terms of engagement that make people willing and able to trust each other enough to launch and sustain society as a cooperative venture. Among separate agents, reciprocity nurtures a solidarity worth wanting, a norm we want to see spreading through a population of players if we care about them.

VI. Toward a Real Ideal: Compliance

The concept of compliance has a confusing place in the literature on ideal theory, as every consumer of the literature knows. Ideal theory is not to be simply equated with the assumption of full compliance, but the latter is the literature's lightning rod, attracting much of the critical attention. We sense something wrong with assuming full compliance. We also intuitively see

that if we were to achieve perfect justice, one feature of that achievement would be that everyone would be in compliance with its requirements.

Between those intuitive understandings, there is room to observe that if compliance is ideal, then taking compliance for granted is not. Full compliance is a hard-won achievement. It is achieved only in certain settings. That means that if there is a realistic ideal, it will be one of those settings. That is, a realistic ideal is a setting in which full compliance is realistically achievable.

Against Rawls, Cohen objects:

> if we assume, following Rawls, that individuals are motivated to comply with justice, then the need to trade off equality and well-being disappears. It only arises in the first place because talented people demand incentive payments to become more productive. But people who are motivated to realize justice fully would not demand incentive payments but rather increase productivity without them.
> (as Cohen 2008 is paraphrased by Hamlin and Stemplowska 2012: 57).

Cohen has a point, but what he misses here is that Rawlsian contractors accept a tougher assignment: they are contracting for a social world populated by agents *other than themselves*. Bargainers cannot stipulate that human psychology is something other than what it is. For bargainers to know human psychology, as Rawls assumes, is for bargainers to know that psychologies of citizens at large are exactly what they are. (See also James 2012: Chap. 4.)

Human psychology being what it is, whether people respond compliantly to what we ask depends on what we ask. Suppose we aim to sell coffee pots. We ask what a profit-maximizing price would be. Are customers disposed to pay what we ask? Obviously, it depends on what we ask. But suppose we imagine that setting aside this pivotal fact lets us focus on the "main" question of what we should charge for our product. Thus liberated from "distracting details," we have an open road to the desired conclusion that customers *ideally* will give us everything they have. What can go wrong?

Crucially, whether customers will give everything they have for a coffee pot determines, not whether asking for everything they have is *feasible,* but whether asking for everything they have is *desirable*. In other words, whether a *bizarrely high* price is the ideal price depends on whether customers are *bizarrely* compliant in the first place. It will not do to argue that in ideal theory agents *are* bizarrely compliant and therefore asking a bizarrely high price *is* ideal. That does not follow (and is not true). All that follows is that if the posited agents were sufficiently unlike real people in a particular way, then asking a bizarrely price *would* be ideal. Theorists

who doubt that ideal prices are sensitive to motivational features of agents do not know what makes ideal prices ideal. An ideal price does not set aside the fact that the point of setting prices is to sell a product. To be an ideal price is to be an ideal response to customers as the separate agents they are. As with ideal prices, so with ideals in general.

If strict compliance is part of an ideally just society's essence, that *precludes* rather than mandates taking strict compliance as given. If indeed "an important feature of a conception of justice is that it should generate its own support" (Rawls 1999: 119)—then a serious investigator will not *stipulate* that a conception has that pivotal feature (when in the real world it patently does not). A serious investigator *checks*.[21]

If instead we assume ideal bargainers not only take for granted their own compliance, but also the compliance of *others*, then we are *failing to check* whether an alleged ideal can generate its own support. Rather, we are imagining what it would be like not to need to check— not to have a political problem. Once we cross that line, we are no longer stipulating simply that ideal bargainers are honest. We are now stipulating that bargainers do not grasp a fundamental political truth: namely, we are not alone. We live in a strategic world of separate agents who decide for themselves.

This is a fundamental truth, not a distracting detail. To set it aside is to ignore the human condition. This fundamental feature conditions nearly everything we have good reason to do— or theorize about—as social animals. What we need from each other, more than anything, is to be able to trust each other. Therefore, our core *political* imperative is to create conditions under which we can afford to trust each other. We are supposed to be theorizing about how to form a community, keep it together, and make it worth keeping together. For social animals, most of what we are capable of as cooperators depends on trust. Trust is a precarious, ongoing achievement. We are failing to imagine what *is* ideal when we imagine what *would* be ideal if only everyone else's reliability were unconditionally guaranteed.

Let's not confuse this with talking about policy as opposed to theory, as Cohen seems to do. To say political theory is theory about what holds communities together and makes them worth holding together is not to propose a policy; it is to identify political theory's subject matter.

Every conception has its own characteristic compliance problem in any given setting. We cannot stipulate that a conception has a compliance problem other than the one it actually

will have in a particular setting. When we choose a package of principles to build into a basic structure of society, we *choose* whichever compliance problem goes with that package.

We choose it! To set aside our chosen compliance problem as a detail best ignored is to set aside *the nature of what we are choosing* as a detail best ignored.

VII. CONFLICT AND CONCLUSION

The topic of ideal and nonideal theory is seductive. The urge to write on this topic is like the urge to gamble: we can't win, so there is no natural satisfaction and no way to get it out of our system. Still, it is time to take stock.

I *accept* that ideal theory can be done well. I *defend* ideal theory in two ways.

First, ideal theory simplifies. All maps and all theories idealize in the sense of leaving out details. To theorize is to set aside details, and to set aside details is to exercise judgment. Even if you do not beg the question, critics will think you did. When you simplify, beware of the impulse to simplify *with prejudice* by setting aside, as 'distractions', truths that reveal that your solution is not ideal, that it will not work, and that you still have a lot to learn. Simplifying is risky: whether a detail is crucial or merely a distraction will depend on the purposes of particular users.

So, are we illuminating real problems or making them invisible? Setting aside distracting details is not a mistake, but setting aside the problem is (O'Neill 1987). Moreover, knowing the difference is an achievement. There is no formula. It takes insight. Authors choose which problem to try to solve. Readers decide whether a chosen problem is important, and when authors try too hard to contrive something resembling a solution they end up changing the problem to something not worth solving.

Relatedly, any decent theory simplifies in order to be useful, thereby leaving itself open to counterexamples. Theories are tools, and one thing we cannot build into useful tools is a guarantee that they could never be misused by evil geniuses. Still, although much can go wrong when tools are useful enough that it becomes possible to misuse them, simplifying per se is not a mistake.

Second, ideal theory articulates objects of aspiration. This too is legitimate in principle. Rawls was right that ideal and nonideal theories are complements rather than rivals.

Articulating objects of aspiration is not a mistake. But there is a difference between aspiration and worthy aspiration. *Calling* something worthy does not make it so. We need theorizing about what makes one aspiration worthier than another. Calling worthiness fact-insensitive does not help. If instead we start with a problem, then our starting point has the potential to discipline our reflection on what to count as a solution. From there, if we develop a view about *what would be ideal*, our ideals can emerge from our theorizing as warranted conclusions about the real world rather than as unwarranted premises.

Taking compliance as given—imagining that people do not decide for themselves whether to comply with what we call justice—is a way of setting aside rather than working on the defining problem of politics: namely, that people choose for themselves, yet still manage under favorable conditions to launch cooperative ventures for mutual advantage. Crucially, favorable conditions are those that induce compliance with norms of mutual advantage. But seriously studying what induces compliance does not start by taking compliance as given. If anything needs to be set aside and treated as a mere distraction from work worth doing, it is visions of how well a system would work but for the recalcitrant reality of human beings.

---

[1] I gesture here at a contrast between my view and David Estlund's. Estlund says,

> People could be good, they just aren't. Their failures are avoidable and blameworthy, but they are also entirely to be expected as a matter of fact. So far, there is no discernible defect in the theory, I believe. For all we have said, the standards to which it holds people might be sound and true. The fact that people will not live up to them even though they could is a defect of the people, not of the theory. (Estlund 2008: 264)

Estlund's insight is that it takes two to make a motivational failure. The bare fact of people responding badly to a standard does not *entail* that the standard is faulty. Of course we may predict that students will fail my exam without blaming my exam. Still, responsible reflection on a predictably bad outcome does not simply make that point and then stop. That students predictably misread double negations is not a defect in my exam, but littering my exam with double negations is. Once I know this about students, the implication is clear: My exam is bad and I need to fix it. It might be ideal for ideal students, but it is not ideal for mine. Or, suppose it is entirely to be expected that insects will evolve resistance to what seemed like an ideal insecticide. Does that count against insects or against the insecticide? Should we infer that the insects are defective? To be sure, we have a choice, but the only honest answer is that we never had an ideal insecticide. Further, whether it was ideal was never a question to answer by envisioning counterfactual conditions under which the insecticide would have worked.

[2] Cohen embraces this possibility (2008: 271–2 or 302–4). To Cohen, saying humanitarian values sometimes trump egalitarian values is (inexplicably) not the same as saying the normativity of egalitarian values is fact-sensitive.

[3] See also Ronzoni 2012.

[4] See Charles Larmore 2013 and Part 1 of Schmidtz 2006.

[5] This is the subtitle of Hume's *Treatise of Human Nature*.

[6] Williams had no interest in underwriting the status quo. To say we need people to form mutual expectations and rely on them enough to form a community is not to valorize any particular order, including the status quo.

[7] Rawls saw that question coming, but it is hard to imagine a worse answer than to say, "rational contractors regard the producers they represent as bringing nothing to the table but morally arbitrary accidents of the genetic and

social lottery to which they have no claim. Said producers, having been represented so capably, will strictly comply with whatever we condescend to distribute to them."

[8] See Part 4 of Schmidtz 2006.

[9] For a survey of choices that need to be made in order to bring ideals of equality to bear on institutional practice, see Brighouse and Swift 2014. See also Dennis McKerlie 1989.

[10] See Robeyns (2008: 358). See also Part 5 of Schmidtz 2006.

[11] On the imperfect duty of reciprocity, see Part 3 of Schmidtz 2006.

[12] Axelrod 1984. I allow that the strategic case is not the only cases. There are nonstrategic situations—early child rearing, say—where children taking full advantage of a free ride is the relationship's point. Thus, it is a mistake to assume *ideal* compliance and *partial* compliance are mutually exclusive by definition. When strict compliance is the ideal (the compliance most worthy of aspiration) it is not settled by stipulation. It is interesting that Rawls, for no apparent reason, comes close to denying this when he stipulates that, "Principles are to be chosen in view of the consequences of everyone's complying with them" (1999: 114).

[13] Reciprocity interpreted along Axelrod lines involves responding to cooperation with cooperation and to defection with defection in an iterated *Prisoner's Dilemma*, thereby making it pay to be a cooperator.

[14] Stemplowska and Swift (2012: 387) say this, making a point about unrealizable ideals in general rather than unconditional giving in particular.

[15] Valentini (2012: 659) says this, likewise making a point about unrealizable ideals in general.

[16] As Rawls warned, "If a conception of justice is unlikely to generate its own support, or lacks stability, this fact must not be overlooked" (1999: 125). Section 15.6 returns to this point.

[17] I employ the phrase "all people together" with a nod to David Estlund's view that a theory can be "hopeless" if too few will comply to make it work as a framework for cooperation, yet still "prescribe action in a certain way" insofar as it "counsels all people, together, to behave differently" (Estlund 2008, 266).

[18] I thank Mario Juarez for the thought.

[19] I thank Robert Van't Hoff for the thought.

[20] I thank Max Kramer for helping me to clarify the point.

[21] Rawls's views about "strains of commitment" likewise suggest that he would be comfortable with the idea that compliance is an achievement that cannot be taken for granted, and thus that his conception of the nature of defensible ideals and his concerns about stability were evolving in the direction of the view defended here.