Res Publica (2017) 23:1–22
DOI 10.1007/s11158-015-9294-x

# The Gauthier Contract: Applicable or Not?

**Jeremy Neill**[1]

Published online: 27 October 2015


**Abstract** In a 2013 article, David Gauthier noted upon the twenty-fifth anniversary of the publication of *Morals by Agreement* that his contractarian approach to morality had found a niche among 'some of those who remain unpersuaded by either Kantianism or utilitarianism'. In this article I will focus on Pareto optimization and I will argue that the Gauthier contract, even in spite of the article's revisions, is still less useful for consultation purposes than Gauthier is assuming. To highlight the conceptual distance that I think separates the Gauthier contract from real-world circumstances—a separation that is greater than Gauthier supposes—I will focus on the problem of 'civic defection'. The point is not for us to repudiate as such the deliverances of the Gauthier model, but simply to use caution in our deployment of the model, and, in order to access the full extent of the model's offerings, to resist the temptation to deploy it in the service of purposes which it is not capable of accomplishing.

**Keywords** Gauthier · Ethics · Rational choice · Contract · Ideal theory

## Introduction

In his 2013 *Ethics* article, David Gauthier noted upon the twenty-fifth anniversary of *Morals by Agreement* that his contractarian approach to morality had found a niche among 'some of those who remain unpersuaded by either Kantianism or utilitarianism', and he updated that approach by substituting 'Pareto optimization' for *Morals by Agreement*'s 'constrained maximization', and likewise 'maximin relative benefit' for 'maximin proportionate gain' (Gauthier 2013, pp. 601–624). In this article I will focus

✉ Jeremy Neill
jneill@hbu.edu

[1] Department of Philosophy, Houston Baptist University, Houston, TX 77074, USA

on Gauthier's new strategy of Pareto optimization, and, arguing that the Gauthier contract is still less useful for consultation purposes than he is assuming, I will in the 'Introduction' section unpack Pareto optimization's details. In the 'Pareto Optimality and the Revised Gauthier Contract' section I will assert that the Gauthier contract's empirical connections are still too minimal for it to offer the consultation help that it would need to offer if it had the real-world implications that he thinks, and in the 'Civic Defection: Does the Gauthier Contract Offer Consultation Help?' section there after I will anticipate responses.

## Pareto Optimality and the Revised Gauthier Contract

In the recent 'Twenty-Five On' article, Gauthier continues in broad outline the project begun in *Morals by Agreement*—to conceive of morality as a system of rules which rational agents would endorse in an initial situation—and he proposes that 'we think of a person as an autonomous cooperator who may demand of society that it defend the normative claims it makes upon him, that it show him he must rationally accept them' (Gauthier 2013, p. 617). The question which he thinks that we need to highlight is 'if, per impossible, you were to be choosing, together with your fellow humans, the terms on which you would interact with them, then what terms would you accept' (Gauthier 2013, p. 618)? In answering this question in 'Twenty-Five On' Gauthier again takes himself to be offering a rational justification of morality and a resolution to the conflict between morality and the pursuit by agents of their personal interests. In his view it is possible for agents to optimize their interests and also the interests of their neighbors by imagining themselves within a hypothetical contract, and then by laying out the conditions for a rational and voluntary moral agreement among persons who are disposed to coordinate their actions alongside their fellows. For Gauthier (1986), amoral but rational agents would elect to cooperate and would in fact dispose themselves to cooperate with his system of normative constraints.

In 'Twenty-Five On' Gauthier once again assumes that the initial situation contractors are rational, able to negotiate in ways that are not restricted by moral preconceptions, and the possessors of realistic capacities. They face a decision-making scenario in which it would for each be advantageous to negotiate in a way that would obtain for her a larger basket of benefits. Even though each would have reasons to secure an outcome that is personally optimal—whether cooperation or defection—in view of the decision-making outcome(s) that she anticipates the others as choosing, nevertheless, as in *Morals by Agreement*, in 'Twenty-Five On' Gauthier once again asserts that rational agents would not privilege a direct maximization strategy since it would not in his view be irrational for them sometimes to take less than they might otherwise get if they were being direct maximizers.[1] Nor does Gauthier think that his rational agents would need in the

[1] '[o]n the maximizing conception it is not interests in the self, that take oneself as object, but interests of the self, held by oneself as subject, that provide the basis for rational choice and action' (Gauthier 1986, p. 7).

initial situation to pursue a common good in order to choose morality. Again, Gauthier also now rejects a third solution that he at one time endorsed in *Morals* when he described his contract in terms of constrained maximization, in the belief that 'rational agents would constrain their pursuit of their own greatest utility in order to bring about mutually advantageous Pareto-optimal outcomes, when straightforward maximization, calling for best reply strategies, would yield only nonoptimal returns' (Gauthier 2013, p. 608).[2] In 'Twenty-Five On', Gauthier lays aside his one-time belief that we ought to understand the behaviors of rational agents in terms of maximization. Instead he now thinks that we should characterize their actions as being predicated on an 'agreed-basis engagement', by which he means that they would take their reasons for action from considerations of fair Pareto-optimality—provided, of course, that their neighbors are similarly cooperative. Pareto-optimality for Gauthier is a measure of efficiency and a cooperative setup in which no further Pareto improvements can be made because no one can become better off without making at least one other person worse off.[3] Thus Gauthier's main shift between *Morals* and 'Twenty-Five On' is that now instead of characterizing the strategies of his rational agents in terms of constrained maximization he thinks that we ought rather to conceive of them in terms of Pareto optimization and to recognize that it would not be rational for each group member to take less than, acting together, each could get.

Pareto optimization is in one sense intended by Gauthier to be like constrained maximization in that it is supposed to produce rational agents who, being aware of their own capacities and also of the concrete costs of their views, would choose morality: a system of impartial constraints, upheld by their peers, upon their personal interests (Gauthier 1986, p. 7, 9).[4] It is also supposed to be like constrained maximization in that it is intended to suggest that cooperators would select morality solely as a subset of the principles of rational choice, and thus that they would not derive it from pre-conceptual moral sources. But Pareto optimality is supposed to differ from constrained maximization in that Gauthier now thinks that in order to obtain Pareto-optimal benefits the cooperators would need to recognize the importance of cooperation alongside each other in accordance with a common set of directives—provided that their neighbors also accept these directives. They would need to recognize that they ought to seek to optimize their personal utility by adhering on an agreed-basis to those common directives that would optimize what they, society members as a whole, would get in the final outcome and thus that would restrict their interference in the utility optimization efforts of others. So for Gauthier the goal of Pareto-optimal agreement, rather than being to yield a negotiation process in which the participants receive their moral benefits in accordance with their relative bargaining power—that is, to trace their moral

[2] To be a utility maximizer is not necessarily to maximize one's non-tuistic preferences. I follow Christopher Morris: contractarians can defend the view that utility-maximization involves the maximization of agent-relative coherent preference. Such a conception of maximization allows for both non-tuistic and tuistic preferences.

[3] Pareto optimality is a way of measuring social efficiency: an outcome is Pareto optimal if no other outcome would make every participant at least as well off and at least one participant strictly better off.

[4] Also, Gauthier (1997, pp. 132–148) and Gauthier (1993, pp. 24–40).

benefits to their respective threat capacities—is instead to obtain agreement among the negotiators on a common set of principles, in accordance with which they could then resolve their various interest conflicts. The most prominent of the Pareto-optimal directives that Gauthier highlights is the principle of maximin proportionate benefit or gain—a cooperative standard from which each could gain and which Gauthier considers to be efficient and fair because it 'compares outcomes in terms of the proportion of potential cooperative gain obtained by each person, and makes the least proportion of potential gain as great as possible' (Gauthier 2013, p. 614). In selecting maximin proportionate benefit the aim of Gauthier's fully rational agents is to treat each other as negotiating equals, so as to fulfill the strict Pareto-optimality condition and to ensure that their personal bargaining power does not influence their cooperation efforts.

Note finally that the level at which Gauthier thinks that his rational agents would be Pareto optimizers is still that of their dispositions to choose—and not that of their individual choices—since because of their conviction that their neighbors would also strive to be moral he thinks that they would dispose themselves to cooperate even when doing so might not always in their individual choices realize the utility of their common directives set. They would simply, as the participants in a Pareto-optimal system in which the point is to abide by common directives, elect to act in accordance with that system and as such to ensure that they would themselves be treated in the ways that they would want to be treated. Thus in Gauthier's view their dispositions would be transformed by their Pareto optimization efforts, since through their negotiations with their neighbors they would become disposed in fact to *want* to comply with their moral agreements—irrespective of whether their choice to be moral would lead in particular situations to utility optimization. As such, within a society—a cooperative venture for mutual fulfillment—the Gauthier of 'Twenty-Five On' still thinks that his rational agents would abide by moral strictures because, having disposed themselves to be Pareto optimizers, they would recognize the rationality of cooperating on the basis of the expectation that their neighbors are likewise cooperating.

## Civic Defection: Does the Gauthier Contract Offer Consultation Help?

In spite of the revisions of 'Twenty-Five On', the Gauthier contract is still less useful for consultation purposes than might at first be thought.[5] To highlight its conceptual distance from real-world circumstances, I will in this section employ the problem of 'civic defection' as a case study of the manner in which real-world persons might seek to derive guidance from it. To be sure, I acknowledge that the civic defection problem applies Gauthier's ideas to a public and political domain that differs importantly from his own, non-contractual initial situation. The first of the two reasons why this application is nevertheless useful is that there are numerous philosophers in the tradition, including Aristotle, who have believed that the moral and political realms are attached, and that a rational justification of morality must in order to have real-world traction be structured and enforced by a

[5] My critiques draw to some extent on Murray (2007) and also Southwood (2010).

corresponding political order (Aristotle 1941, 2001).[6] Without such an order, Gauthier's distinctively non-empirical account of morality's rationality would be for real-world persons of little more than mere educational edification. The second reason why this application of Gauthier is useful is that the Gauthier contract is in its original formulation being pitched abstractly and that certain of its aspects—including especially its impracticality costs—are more profitably illuminated via a real-world case study.

That said, the most comprehensive contemporary documenter of the civic defection problem, Robert Putnam, has used 'social capital' as a metric of citizen engagement and, focusing on the United States, has charted a decline in American civic involvement in the last sixty years: Americans have become less likely to participate in their civic associations, less interested in public discourse, and less ready to incorporate their lives into a political narrative; they are cynical about their institutions, they do not believe that individual involvement is important, and they do not identify with the general welfare.[7] Among their attitudes these days are apathy and a disinterest in politics; among their beliefs are (1) that their efforts as individuals cannot alter the political process; (2) that personal fulfillment is more achievable in non-civic pursuits, and (3) that the public sphere would be just as functional in their absence (Galston 2007, pp. 625–626, 628, 636–638, 639–640).[8] They see defection as being more rational than participation because they think of the public sphere as a setting in which their participation is justifiable as long as it achieves their interests (or, for many, the collectively Pareto-dominant interests), and from which defection is acceptable when its benefits are diminished.[9]

The Putnam studies imply that the American civic defections are not defections from cooperation as such, since, even as Americans today are defecting en masse from their civic institutions, they are still cooperating in a more basic sense with their immediate neighbors. Their defections are *partial defections* because they are still cooperating in local ways that have only limited or indirect civic implications: in their homes and families, their friendships and local workplaces, and their consumer purchases. Thus it is reasonable (if a bit arbitrary) to parse the American moral system into a first stage at which the citizens cooperate with each other normatively as such, and a second stage at which a certain percentage of them participate in the institutions that facilitate their first-stage cooperation. This two-stage conception of the American moral system, though rough, is useful for my purposes because it captures the stabilizing roles of their institutions, the disparities

[6] Without the political domain, how else in a large liberal democracy would morals be enforced?

[7] See Putnam (2000, pp. 31–64, 277–284), Galston (2001a, p. 16), Elshtain (1995, pp. 1–36), Galston (2004, pp. 263–266), Galston (2001a, pp. 217–234), and Dagger (1997, pp. 132–153).

[8] Activities which are being abandoned include voting, demonstrating, lobbying, and running for office. See Putnam (1995, pp. 664–683), Rahn and Transue (1998, pp. 545–565), and Costa and Kahn (2001).

[9] Their partial defections are representable with the preference relation R, where $x$R$y$ means that $x$ is at least as good as $y$. If agents believe that $x$R$y$ and not that $y$R$x$, then for them, $x$P$y$—$x$ is preferred to $y$. The participation of the civic defectors is describable as $y$P$x$, where $y$ is civic defection and $x$ is civic participation. Yet, typically, they also believe that $y$P$z$, where z is a further step – defection from society as such – and $y$ is merely an initial step of civic defection. In selecting both $y$P$x$ and $y$P$z$, they are electing to be first-stage cooperators and second-stage defectors.

in their various collaborative investments, and the resilience of the normative system in the face of their partial defections. Its gist is that second-stage institutions are central to American morality and that for functionality's sake the system requires widespread second-stage participation. In Gauthier's parlance, the Americans in Putnam's studies, in endeavoring to be first-stage cooperators but defecting at stage two and pursuing their interests in other ways, are in their civic affairs—though not in the rest of their lives—free riders who are choosing other strategies than Pareto optimization: whether straightforward maximization, which does more for their personal interests than Pareto optimization, or a Nash-Harsanyi bargaining strategy on which the benefits of institutional participation do not meet their threshold cooperation requirements vis-à-vis the threat advantages of others, or something else.

For the sake of the example, let us suppose that the American moral system grows more functional the more that Americans, collectively, are second-stage participants. To be clear, I am not asserting here that greater system functionality follows by necessity from numerical increases in second-stage citizen participation. But the social science literature *does* suggest that the American civic order increases in functionality the more that second-stage participation increases.[10] If so, then America's widespread civic defections ought to be seen as being a social problem—even if the citizens are still first-stage cooperators—since their second-order institutions can typically do less to protect the first-order moral order when there is mediocre second-stage involvement. Thus, while it is not a moral wrong as such for Americans to be second-stage defectors, nevertheless the empirical evidence is that their collective second-stage participation improves their country's cooperative functionality. Yet, at the same time, even in spite of this historic functionality decline, note that the impact of the civic defectors ought not to be seen as being catastrophic because there are still numerous other Americans who are in spite of the drawbacks still endeavoring to be second-stage cooperators. They are altruists who are rendering the American moral system, such as it is, minimally functional. The thought of many of the civic defectors is that others in the culture are responsible and are choosing institutional participation, and as such that it does not matter whether they are themselves being responsible or not.

The reason why the Gauthier contract appears at first glance to offer consultation help is that the model is most naturally applied in something like the way I have described. That is, in order for Gauthier's Pareto optimality strategy to be a practicable amid modern social conditions, it arguably for reasons of surveillance and enforcement would be most optimally manifested as a two-stage process—with robust second-stage participation. This is an important claim and I am making it because in the background in 'Twenty-Five On' the transparency/translucency distinction of *Morals by Agreement* is still present in the conditional strategies of Gauthier's Pareto optimizers: they only cooperate if they expect their neighbors to cooperate, and they are unwilling to cooperate when they are surrounded by uncertainties because they know that if they are unable to ascertain the cooperative

[10] For evidence that liberal democratic governance is more functional the more that the citizens participate, see Putnam (1993) and Knack (2002).

intentions of their neighbors then their own Pareto-optimizing efforts are unlikely to be profitable.

Now amid our modern social conditions there are many persons who, not having the luxury of the relational proximity of past cultures—in which traditional community ties facilitated cooperation—are in their social interactions strangers, and would in the absence of second-stage institutions struggle with cooperative discernment. This is what suggests to me that the Gauthier contract's behavioral expectations are best realized via second-stage institutions. That is, I think that Gauthier's rational agents would amid our atomistic and individualistic modern cultures require second-stage institutions because they would need such institutions in order to unify and enforce their morality efforts, to publicize their first-stage cooperative intentions, and, ultimately, to discern the cooperative intentions of their peers. Second-order institutions would alert them to each other's aims and would render their first-order moral efforts successful.

In what way do second-stage institutions equip Americans to discern each other's cooperative intentions and how when Americans are interacting with each other as strangers—or at least as persons who do not know each other as well as their ancestors once did in the close-knit communities of the past—can they derive their cooperative cues from their institutions? In part the way in which second-stage institutions render the cooperative intentions of Americans transparent is that their rules and customs alert the citizens to the willingness of their neighbors to be moral. The second-stage institutions publicize the citizens' cooperative intentions and thus make clear the first-stage aims of persons who might otherwise conceal their intentions. Consider the price determination markers of our modern securities markets—our stock and bond prices, currency exchange rates, interest rates and dividend yields. Via advertising and publicity incentives, these markers offer information that often is more trustworthy than what we could otherwise obtain in less regulated settings. The upshot is just that Gauthier's Pareto-optimizing cooperators would amid our modern social conditions need second-stage institutions in order to clarify their cooperative intentions, and as such that the Gauthier model resembles the American moral enterprise and could in theory be a consultation template.

How might Americans look to the Gauthier model for consultation help and how might the civic defectors be convinced once again to participate? The way in which Gauthier in 'Twenty-Five On' convinces his bargainers to cooperate is by telling them that doing so would advance a collectively payoff-dominant outcome in which no one individual could be made better off without making at least one other individual worse off. Since the belief of Gauthier's rational agents that it cannot be rational for each of a group of persons to take less than, acting together, each can get is at least prima facie similar to the partiality orientations of contemporary Americans, perhaps Gauthier could analogously say to the civic defectors that their return to civic life would also advance the collectively payoff-dominant outcome: their common policy goals, their material prosperity, their opportunities to be freely and fairly treated by others, and their relational flourishing. The aim of this argument from Pareto-optimal advancement would be to persuade the civic defectors once again to be civic participants, for the sake of the collectively payoff-

dominant outcome, and in a manner that is analogous to their first-stage efforts to rely upon the moral goodness of their neighbors. Again, assuming that the American moral system typically functions better the more that Americans are endeavoring to be second-stage participants, it seems naturally to follow that institutional involvement would similarly optimize what citizens, the society members as a whole, could get at the level of dispositions to choose (although not, of course, at the level of individual choices). Among the various collectively payoff-dominant interests that Americans could in theory advance through second-stage involvement are their common policy aims, their feelings of personal fulfillment, their friendships, financial gains, ethical stability, and sense of national pride.

Nevertheless, upon closer inspection the Pareto-optimizing argument is less promising than it initially appears, and as such second-stage institutional participation is not as reliable a way for Americans to optimize the collectively payoff-dominant outcome as it might initially seem—even at the level of dispositions to choose. I say this because the trends toward individualism and social atomism are increasingly prevalent in American culture, and together they undermine the ability of persons to achieve, through second-stage institutional involvement, non-negligible advances in the collective payoff-dominant outcome.[11] In fact the great majority of the institutional benefits in America arise in a manner that is not capable of being statistically correlated to individual participation: the enormity, complexity, and lethargy of the institutions are causing the benefits to arise in the same way and with the same regularity whether or not particular individuals are involved. At stage two of the collaborative process, American institutions are offering a cooperation form whose Pareto-optimal institutional benefits are indifferent to individual participation choices: the social security checks are dispersed whether or not a particular citizen chooses to vote, the city roads are repaired whether or not she attends her town forum, and the military defends the well-being of the public whether or not she is aware of their efforts. Thus, second-stage institutional participation is not in practice capable of yielding non-negligible Pareto-optimal benefits improvements for most Americans, and it is reasonable to conclude that many of the citizens who are today still striving to be civic participants are, at least as individuals, no more likely than their defecting counterparts to advance the collective payoff-dominant outcome.

Suppose that the interests that matter the most to the citizens are the policy-making goals that would for them, the society members as a whole, be of greatest benefit. Suppose also that there are three classes of liberal democratic institutions in America: A, B, and C. Class A institutions are extensively associated with governments and are embedded in the society's constitution, in legislative applications of the constitution, and in similar political injunctions. Imaginable in class B are institutions that are less closely associated with governments and that are viewed by the public as autonomous, and perhaps even as having a right to their organizational independence. Class C institutions are spontaneous forums that are wholly unregulated and that pass in and out of existence in accordance with a society's conversational needs.

[11] See Taylor (1985), Crowder (2002), Gans (1991), Grabb et al. (1999), and (Bellah et al. 2007).

Say an ordinary American wanted to achieve the policy aims that would for the society as a whole be of greatest benefit. Would it be rational for her at stage two to seek to optimize the collective Pareto-dominant outcome via institutional participation? Typically, such a strategy would be no more rational than defection. In contemporary America the bureaucracies are vast and the political powers are shared among many thousands or even millions of citizens. Most class A and class B institutions are so large and complex that it is not possible to draw a meaningful statistical correlation between their policies and the individual involvement of their participants. It is rarely feasible amid such institutional vastness for ordinary Americans to make a policy-making difference, of a kind that could do non-negligible things to advance the collectively payoff-dominant outcome. Only the political elites are ever able to make a non-negligible policy-making difference. The vast majority of the citizens—the ordinary folks—can only bring about policy-making payoff-dominant outcomes when large numbers of their fellow citizens support causes that are similar to their own.

The inability of the non-elites to alter their institutional values, and more generally to produce a Pareto-optimal outcome, is at its most obvious in the large class A institutions: in America, Medicare, the Department of Housing and Urban Development, and the FDIC are all sizeable and lethargic, do not change easily in response to individual efforts, and yield their benefits in a way that is statistically unrelated to individual involvement. Neither is it typical for the non-elites to make non-negligible differences in the sizeable class B institutions: universities, hospitals, multi-national corporations, or media networks. Only in the small class B institutions and the more disorganized class C institutions—neighborhood organizations, churches, veterans' organizations, and internet chat rooms—is it possible for the non-elites in non-negligible ways to advance the collectively payoff-dominant outcome. Yet the latter such institutions are often too nebulous and too small for their benefits to be realizable in a way that would optimize the outcome that the society members as a whole could get, in terms of collective payoff-dominance.

Similarly, the other, non-policy Pareto-optimal interests that are capable at least in theory of being promoted through second-stage citizen involvement—higher salaries, better personal relationships, and more coherent communications—also usually are no more capable in practice of being advanced through second-stage participation than they are through second-stage defection.[12] Thus the reason why second-stage Pareto-optimizing participation would not in practice be more rational than defection is that at stage two of the country's cooperation system the institutions are large and impersonal, and render the involvement of individual Americans of merely negligible consequence from a Pareto-optimal perspective.[13]

[12] American Millennials are increasingly viewing the Pareto-optimizing impact of civic defectors as being just as great as the Pareto-optimizing impact of civic participants. Instead of choosing lives of civic service the Millennials are turning increasingly to the private sector, believing that their capacity to promote a Pareto-optimal outcome is just as great through civic defection as it is through second-stage participation. See Duncan (2012), Lapin (2012), and Denz (2012).

[13] The institutional system is an indifferent cooperator and not an unconditional cooperator. It does not cooperate at all times—the roads are not always reliably repaired.

The policy developments that are positive from a Pareto-optimal perspective, and more generally the benefits of second-stage institutions, are for the citizens obtained in a manner that is indifferent to their individual participation choices. So it does not matter from a Pareto-optimal perspective whether they are civic participants or not.

In fact in mainstream class A and B institutions the most that the Pareto-optimality participation argument could promise the citizens is an indirect, slight, or tenuous opportunity to promote a collectively Pareto-dominant outcome. To be sure, in America's small and local class C institutions—neighborhood organizations, hobby societies, and internet chat rooms—it might be possible sometimes to identify statistically-meaningful correlations between Pareto-optimal institutional developments and the inputs of particular persons. Perhaps if such links were robust then they could convince the skeptics that their class C institutional involvement would optimize what they, the society members as a whole, would get in the final outcome. Yet in the great majority of class A and B institutions—the public universities, the public court system, Medicare, HUD, or large private institutions like the AARP, NBC, and the American Bar Association—any effort to show the citizens that their second-stage participation would do more than non-participation to produce collective payoff-dominant outcomes would need to rely on indirect and nebulous influence threads: the ways in which their second-stage participation might bring about a marginal structural improvement here or a slight opinion adjustment there—changes that would, over time and via a causal series, optimize the outcome that the society members as a whole could get. Such threads are unlikely to convince the civic defectors to return to the public sphere, since they are unlikely to clearly and distinctly show that individual participation is doing more than other things for what the society members as a whole could get in the final outcome.

More convincing as demonstrations that second-stage involvement is capable in non-negligible ways of promoting Pareto-dominant outcomes would be studies of smaller and less formal class C institutions which spelled out directly for the citizens the connections between their personal involvement and the development in their institutions of particular policy changes. If such influence threads were shown to be robust, then presumably the citizens could also be convinced that their own second-stage participation could do non-negligible things to promote a collectively Pareto-dominant outcome. Its benefits would include their policy goals, material prosperity, opportunities to be freely and fairly treated by others, relational flourishing, and psychological satisfaction. Yet since in practice these robust threads are unlikely to be forthcoming in the great majority of class A and class B institutions, the Pareto-optimality participation argument probably would not usually be for ordinary Americans a convincing defense of class A or B institutional participation.

Thus what contemporary America offers is a case study which resembles the Gauthier decision-making situation and which illuminates the real-world efficacy of the Gauthier contract. American moral cooperation is a multi-staged affair whose different layers make it reasonable for the citizens to pursue the collectively Pareto-dominant outcome in certain areas of their lives and yet also at the same time to behave in other ways in other areas: to be first-stage cooperators but then perhaps to defect from stage two once they determine that their personal participation is not

doing non-negligible things for the cause of Pareto optimality.[14] The case study suggests in particular that it is not in a dispositive way rational for persons to participate in at least some levels of a multi-stage real-world scenario which meets their basic needs and in which their personal participation fails to enhance system functionality.

Notice that America's distinctive characteristics are what make it so difficult to offer a convincing and Pareto-optimizing second-stage participation strategy. One such characteristic is the system's large size, which renders individual defections unimportant vis-à-vis functionality. Another is the system's inadequate participation enforcement. Again, still another is that numerous altruists are today facilitating the system's functionality, even in the absence of personal benefits. A major reason why the civic defectors still more or less promote the same Pareto-optimal benefits as their participating counterparts is that even in spite of the defections there are still today numerous other Americans who are altruists and second-stage cooperators for reasons that are otherwise than Pareto optimality. The efforts of the altruists—which are not utility optimization efforts because as individuals they are not in fact making a Pareto-optimizing difference—are what in a collective sense (although not an individual one) are rendering the second-stage institutions functional, and are enabling the defectors to ride freely.[15] Of course if my earlier assumptions are correct then at the same time the civic defections are still at least in a collective sense—though not an individual one—doing damage to the American moral system, and as such that system is not producing the collectively payoff-dominant outcome, by Gauthier's standard of maximin proportionate gain, as it ought to be doing.

Presumably, in any circumstance that manifests these particular characteristics—large demographic size, inadequate enforcement, numerous altruists, tiered participation, etc.—a Pareto optimization strategy would face similar struggles. So if readers are troubled by the environmental differences between my own, civic and American case study and the Gauthier contract's initial situation, I suggest that they focus on the scenario's underlying structure—its tiered arrangement in which the second tier is essential to system functionality, but in which individual second-tier defection produces no untoward effects—and not so much on the fact that it is civic or American. The point is still legitimate even if the scenario is just a construct because the case study is repeatable in non-civic settings with similar characteristics. Its upshot is just that Gauthier's Pareto optimization strategy needs in order to have real-world purchase to be able to convince persons that their individual Pareto-optimizing participation efforts could in fact produce a Pareto-optimal result. But if Pareto optimizing participation does not in fact meet these threshold-level efficacy requirements, then such persons could presumably pursue other, different forms of

[14] Suppose it were objected that at America's stage two the decision-making scenario is already moralized because its stage-one moral features are already shaping the citizens, and thus the case study does not resemble Gauthier. I respond that the empirical evidence suggests that the second stage is inseparable from the first, because in a society of strangers who lack relational discernment mechanisms, the choice has to be for both, and not just for one.

[15] Think of the senior citizens who donate their time to public causes, the altruistic polling station operators, the citizens who serve for lesser pay on county juries, and the volunteer political campaign workers.

cooperation. The inability of real-world Americans, through second-stage institutional involvement, non-negligibly to advance the collective Pareto-dominant outcome suggests that the efficacy obstacles are just too great for second-stage Pareto-optimizing participation to be rational in any dispositive way. Just as rational is for them to pursue other second-stage strategies, and to defect from their institutions on the basis of their knowledge that their capacity to optimize their own utility vectors and the utility vectors of their neighbors is unchanged by their non-participation. Just as rational is a strategy of partial participation and partial defection—to follow a Nash bargaining strategy, as it were, and to choose institutional defection by allowing the Nash strategy's equilibrium condition to trump the Pareto-optimal strategy's egalitarian condition.

The argument also suggests that Gauthier's contractual revisions in 'Twenty-Five On' have still not solved the free-rider problem amid modern, real-world settings. Instead they have transformed that problem into something that is different from what it once was, vis-à-vis his first contract.[16] In particular, Gauthier's revisions have made the free-rider problem no longer a dilemma of self-interest, in which persons have reasons to defect for the sake of personal gain, but now one of efficacy, in which persons have reasons to defect because their cooperation would not now make a difference vis-à-vis Pareto optimality. Free-riding might actually under the new Gauthier contract be a more intractable difficulty because now the issue is not as much to show that cooperation is self-interested—the free-riders are not now necessarily driven by self-interest—but rather that cooperation is non-negligibly constructive.

In sum, the revised Gauthier contract is still unconvincing as a response to a standard practical case study and thus it probably is less applicable amid our modern social conditions than it might initially seem. At the very least the contract does not solve the free rider problem in the way that Gauthier intends and it is unlikely to convince the defectors to reverse their trajectories.[17] Gauthier's claim that morality results from Pareto-optimizing rationality is unhelpful for consultation purposes in a tiered real-world scenario which is large and in which some persons are willing to be second-stage altruists.

[16] Civic defection of course has been well-traveled in the literature; here I am just saying it is not capable of being solved via a Pareto-optimization strategy. For instance, Brian Barry has argued that Gauthier's contractarianism collapses because it fails to provide a satisfactory answer to the question of why agents should obey the rules if at some time some other action path offers a better prospect of promoting their conception of the good (Barry 1995, p. 34). Examples of works that have highlighted the free-rider problem in a general-purpose civic context include Schumpeter (2008), and Arneson (2009).

[17] Civic defection is just one of numerous different real-world problems that could potentially be referenced to demonstrate the inability of the Gauthier contract to offer consultation help. Others are the tendency of persons to change their minds in the course of their cooperative efforts; the inconsistencies of their behaviors in crises; their inability to discern each other's intentions with the certainty that Gauthier anticipates; and their propensity to deceive each other.

## A Potential Gauthier Response

One way in which a Gauthier defender might reply is by saying that the civic defectors are irrational and that if they were not then they would recognize that Pareto-optimal advances occur whenever the citizens, collectively, are second-stage participants. Say, if Americans were to be educated into genuine Pareto optimality—a state of affairs in which it is impossible to improve any one party's position without making another party worse off—then, taking up its 'we' viewpoint and not their old 'I' viewpoint, they would realize that their collective participation, at least, does make a non-negligible difference. Here I agree that if a Pareto-optimality strategy were an idea that so thoroughly subordinated individuality to the collective that the former were no longer even psychologically present in the decision-making process, then no non-collective considerations could be entertained and participation would be rational. That is, participation would be rational if in taking up the Pareto-optimizing perspective persons were to enter into a state of irreducibly complex and collective decision-making, manifesting sentiments, opinions, and choices that are not traceable to individuals.

But does a pile of sand ever become so numerically and conceptually substantial that it acquires a life of its own and is no longer conceivable in terms of its parts? If this were to be our definition then in my view we would, instead of continuing to interact with the game-theoretic idea of Pareto optimality, find ourselves working rather with the much older and more general-purpose idea of the 'common good'—which Gauthier explicitly repudiates as a way of interpreting Pareto optimality: 'rational cooperators need not seek a collective or substantively common good. Each is concerned to realize his own good, as expressed by his utility function….rational cooperation yields mutual fulfillment; it need not afford common or collective benefit' (Gauthier 2013, p. 609). My own view—shared by Gauthier ally Edward McClennen—is that Pareto optimality is not a state in which the collective is irreducibly complex, but instead that it means that the individual is embedded within the Pareto optimality idea and that he finds his place or fulfillment *through* the collective's cooperation decisions.[18] Pareto optimality is a condition of individual rationality, and not one of collective rationality—and thus a condition of individual, first-personal decisions—so that within the decision-making of the collective's members there still is an idea of individuality. The group is an assembly of persons who each are seeking mutual fulfillment, and thus the idea of the individual agent is still at least a live psychological consideration (i.e. the individual must pass through the 'I' in order to conceive of the 'we'), and thus also it is legitimate for the individual agent to take into his decision-making process—the 'we' decision that promotes the mutual fulfillment of the members of the group—the fact of his own insignificant institutional participation.

The paradox that I am highlighting is just that three things are simultaneously possible in the American case study: the individual employs Pareto optimizing

[18] For McClennen, 'the failure of any arrangement to generate a Strictly Pareto-Optimal outcome means that the arrangement in question fails to meet a condition of individual rationality' (McClennen 2010, p. 531). Also, see Roth (1977, pp. 64-65).

reasoning and seeks the mutual fulfillment of the system's 'we'; he recognizes that the system redounds benefits to everyone in a way that is indifferent to his personal participation; and he realizes also that his defection is justifiable since his participation does not non-negligibly contribute to the mutual fulfillment of the 'we.' Note that he does not have to find justification for his defection on the basis of self-interested desire maximization, but rather that he can find justification in reasons that are consistent with a Pareto-optimizing game theoretic.[19] In a system in which there is not for ordinary persons a non-negligible correlation between Pareto-optimal gains and their second-stage institutional participation, and which fulfills the other operational dynamics that I have described, participation in at least the second of its stages is not rational in any dispositive way. Just as rational is for real-world persons to choose defection on the basis of their realization that their individual second-stage participation does not non-negligibly promote Pareto optimization.

Note likewise that it also is not dispositive, amid a paucity of evidence that their second-stage participation has non-negligible significance, for individuals to assume via faith that their personal participation in fact is Pareto optimizing—it simply does not appear to be possible for our social science studies to produce statistical correlations in most Class A and Class B institutions that suggest that our individual participation makes a difference. The real-world efficacy deficits of the revised Gauthier contract suggest that second-stage defection in fact is just as reasonable for most citizens, since they must by virtue of Pareto-optimal reasoning make their participation decisions with an awareness of their individuality, and since there are, after all, numerous other things that they could be doing that are personally or relationally profitable, and that are neutral vis-à-vis the cause of Pareto-optimal morality (McClennen 2010).

More generally, since Pareto-optimizing reasoning does not appear to produce a dispositive cooperation result in a real-world situation of this kind, my point is just that the candidate agents—Americans (or whoever, if the case study were to be varied)—could also employ other forms of reasoning as ways of navigating their cooperation options. Individuals are free in such a scenario to pursue other cooperation strategies, on the basis of their awareness of the participatory impoverishment of Gauthier's approach.

## The Abstractions of the Gauthier Contract: Too Great for Consultation Purposes

From these reflections there appear to be two ways—descriptive and prescriptive—in which the Gauthier contract is more abstracted than Gauthier thinks, and too abstracted also to offer the consultation guidance that Gauthier anticipates. It is first of all more descriptively abstracted because of the differences between Gauthier's

[19] Note that while defection is just as rational as participation (if not more so, because of the agent's other considerations), it too is not decisively rational because the agent might have an individual cocktail of non-Pareto-optimizing considerations that suggest participation.

Pareto-optimizing reasoning and the reasoning of real-world cooperators. In particular, via the Pareto optimization route Gauthier seems to be assuming a number of morally-laden things about his contractors—including especially their reasonableness and their willingness to curtail their personal interests for the sake of the greater collective interests that they could otherwise obtain as the members of a group. But here Gauthier's Pareto-optimizing reasoning is reducing the multi-faceted motives of real-world agents to a single motivational stream: a motive of agreed-basis utility optimization that is experienced by free and equal agents, and that Gauthier thinks would orient them toward Pareto optimization. Consider, as evidence of the implausibility of assuming that all—or even many—Americans are likely to reason in this way, the social psychology research which has arisen out of the 1990s prisoner's dilemmas exchanges. The thinking of the real-world cooperators who have been highlighted by this research has been designated by the social psychology researchers as their 'social value orientation' (SVO), and it has been described as having a scalar value.[20] The SVO research suggests that the utility optimization that Gauthier highlights is just one of a number of different motives that inform the cooperative efforts of real-world agents. The great variety today in the cooperation motives of real-world persons suggests that their motives are much of the time tuistic, that they typically are mixed in ways that are different from the motives of Gauthier's rational agents, and that as such they are difficult for a theorist like Gauthier to parse from an idealized perspective.[21] Thus although when Gauthier says that his bargainers would cooperate for a collective payoff-dominant outcome he is positing a utility optimization motive that might seem at least prima facie to be similar to the motives of real-world persons, nevertheless he also is at the same time simplifying real-world motives to an extent that is perhaps too great for him then to be able to offer pertinent insights.

The descriptive abstractions of the Gauthier contract are also notable in Gauthier's assumption of a rationality that is more circumspect than standard real-world rationalities. In requiring his Pareto-optimizing contractors to look at their negotiations from a third-personal perspective (albeit one in which the 'I' is present in the background), Gauthier is assuming that negotiations take place from a perspective that is otherwise than ordinary first-personal thinking, and as such he is underrepresenting the experience of social competition—since, after all, it is doubtful whether contractors can enter into meaningful negotiations if they are not actually seeking to advance their own interests at the expense of their neighbors. Thus the Gauthier contract does differ from real-world cooperative dynamics, in which we typically do negotiate, and in a manner that is less impartial than Gauthier supposes.

To be sure, in characterizing the Gauthier contract as being more descriptively abstracted than he thinks, I certainly acknowledge that Gauthier is trying,

[20] See Bogaert et al. (2008), Also, Murphy et al. (2011). Another good overview of social value orientation and the social dilemmas literature is Van Lange et al. (2013); and also Henrich et al. (2005). There is likewise literature on the neurodynamics of prosocial behavior (Declerck et al. 2013).

[21] This point has already to some extent been made, although in a different way and not in a manner that is backed with recent social science studies, by Hubin (1991); also on point is McCracken and Shaw (1995); a response is available in Dimock (1999); again, see Morris and Ripstein (2001).

consciously and for the sake of analytical clarity, to separate his contract from real-world circumstances. Gauthier is not spelling out the blow-by-blow connections between his idealized agents' cooperative efforts and the real-world scenarios which he intends to influence; nor is he trying to list the steps that his idealized agents would need to take in order eventually to apply their agreement. Rather than making the opinions, habits, and motives of his idealized contractors one-to-one reflections of the opinions, habits, or motives of real-world persons, Gauthier's aim instead is merely to demonstrate the rationality of moral constraints for aspiring utility optimizers. Gauthier posits the contract primarily in order to derive a rational basis for morality, and, as such, to trace morality to a non-moral explanation.

At the same time, however, I consider my criticisms of the contract's applicability still to be pertinent because so many passages in *Morals* and 'Twenty-Five On' suggest that Gauthier intends for his initial situation to be practicable, for the psychologies of his agents to be realistic, and for his analysis to be relevant for real-world agents. For instance, Gauthier at one point in the *Ethics* article says that 'we are of course far from realizing a society that would pass the contractarian test, but we can survey the gap between the actual world and the ideal and recognize the steps that have been taken to close it, as well as some of the steps that need to be taken' (Gauthier 2013, p. 620; Gauthier 1991, pp. 15–30). This and other key sections of *Morals*, 'Twenty-Five On', and other writings in which Gauthier explores the psychologies of his contractors are intended to be exercises that illuminate our understanding of human cooperation, and, ultimately, that impact ordinary persons (Gauthier 2013, pp. 613, 620).

The second of the ways in which the revised Gauthier contract is more abstracted than he believes is describable in terms of 'prescription', a metric I will use for measuring theory efficacy. Let a particular moral theory be 'prescriptive' if it offers meaningful consultation help to real-world agents, and in particular provides agents with tangible insights into their moral situations and the ways in which they could improve such situations. The criterion for rating the consultation efficacy of a particular prescriptive theory is its ability to resolve the social and moral problems of real-world persons—in this case, the civic defection problem. Thus for the sake of the argument let the opinions of real-world persons be decisive in determining whether Gauthier's Pareto-optimal strategy is as useful for consultation purposes as he supposes.

With the 'prescriptive' idea in mind, consider again the Pareto optimization strategy and the civic defection problem. For Gauthier, a Pareto optimizer cooperates in ways that would, if followed by all, yield collective Pareto-dominant outcomes. In practice she does in fact cooperate should she expect cooperation to achieve such outcomes. For Gauthier, Pareto optimization does not demand from his negotiators an unfair moral contribution and, as such, he thinks that aspiring Pareto optimizers can establish alongside each other a workable moral system.[22]

[22] This conviction is a reiteration of what he said earlier in *Morals*: '[a] community in which most individuals are disposed to comply with fair and optimal agreements and practices, and so to base their actions on joint cooperative strategies, will be self-sustaining' (Gauthier 1986, p. 182).

But once the Pareto optimization strategy is compared to the SVO research's potpourri of real-world cooperative principles and situations, it becomes evident that the motives of real-world persons are justifiably multi-faceted and that their actual cooperative principles ought usually to be more often haphazard and situational than a strategy of Pareto optimization. Consider the way in which it is sometimes more rational from a Pareto-optimizing perspective for someone to defect when she is interacting with unconditional cooperators (i.e. persons who cooperate at all times), or indifferent cooperators (i.e. persons whose cooperation patterns are independent of her personal choices), for the reasons I have listed. Since in many such situations it would not make a non-negligible difference whether she endeavors to cooperate or not, cooperation presumably would not meet the threshold standard of efficacy that it would need to meet in order in any dispositive way to be rational. It likewise would presumably be just as rational for those real- world persons who do not primarily desire to achieve their personal interests to endeavor still to be altruistic when doing so might not always promote a collectively Pareto-dominant outcome.

There are numerous other such cooperative principles which, given the real-world efficacy deficit of a Pareto-optimization strategy, might at different times and in different places be more rational than Pareto optimization: in one area of life it might be rational for persons to cooperate regardless of what others are doing; in another, to cooperate only if others are being unconditional cooperators and are cooperating regardless of what they are doing; in another, to cooperate except when they think that their defection would yield extraordinary rewards; in still another, to cooperate except when their intentions might be especially hard for others to discern.[23] More concretely, perhaps more rational than Pareto optimality would be for real-world agents to be reserved maximizers with respect to their tax obligations, but at the same time in their immediate families to cooperate unconditionally. Again, it might be rational for them to comply with their university's testing policies only when they are convinced that violations would not yield extraordinary profits for their neighbors, and at the same time to cooperate with their music copyright laws on the basis of a wholly different criterion—one which produces defection except, say, in situations in which violations would be easily detectable.

Gauthier, of course, knows that the motives of real-world persons often are otherwise than the motives that inform his Pareto-optimization strategy. Certainly the psychological fact that real-world people have a variety of reasons for cooperating does not make the revised Gauthier contract wrong, as such, as a defense of rationality-based morality. But it does make that contract less relevant than it could otherwise be and it does suggest that Gauthier might be underestimating the distance between his model and real-world cooperation practices. Gauthier's confidence in the ability of his Pareto-optimization strategy to facilitate cooperation, and more generally his belief in Pareto-optimization's applicability, are upstream stances that enhance his explanatory capabilities but that later, at a downstream stage, abstract his model from real-world cooperation. On the ground,

[23] See Biel et al. (2008).

real-world cooperation is just too diverse and too multi-faceted for Gauthier accurately to parse it in accordance with a single cooperative explanation.

## What are the Purposes of the Gauthier Contract?

Now in the face of the inapplicability criticism that I have derived from the American case study, Gauthier could double-down and reiterate that the descriptive detachment of his contract is in fact the point and that his contract is not as such supposed to offer any one-to-one representation of the motives and relational principles of real-world persons.[24] But if Gauthier were to respond that his model is nothing more than a mere rational justification, and that it is not as such supposed to have tangible moral implications, then he would be introducing other and more incorrigible difficulties—such as, first, undermining some key statements in 'Twenty-Five On', and, second, rendering himself vulnerable to the charge of prescriptive insignificance.

As for the first difficulty, in 'Twenty-Five On' and earlier in *Morals* Gauthier does make it clear that he intends for his contract to be a real-world consultation device.[25] As he says in a 1991 article:

> Actual moral principles are not in general those to which we should have agreed in a fully rational bargain, but it is reasonable to adhere to them in so far as they offer a reasonable approximation to ideal principles. We may defend actual moral principles by reference to ideal co-operative arrangements, and the closer the principles fit, the stronger the defense (Gauthier 1991, p. 28).

For Gauthier, the functionality of real-world cooperative forms ought to be measured against an idealized negotiations template.

This intention is also evidenced by Gauthier's confidence in 'Twenty-Five On' in his own reflective advantages and real-world insights (Gauthier 2013, p. 620; Gauthier 1991, pp. 15–30). Gauthier seems in particular to assume that it would be natural for real-world cooperators to consult his contract—by virtue of the fact that any society that 'claims that its practices and edicts give its members reasons to act must satisfy the principle of maximin proportionate gain' (Gauthier 2013, p. 613).[26] Elsewhere he says that:

[24] Such a response is certainly consistent with Gauthier's sentiments in chapter six of *Morals*, where he suggests that real-world persons act for motives and principles that are otherwise than the motives and principles of his ideal bargainers.

[25] This is true, I think, even though I grant that he does not see its normative deliverances as being one-to-one replacements for real-world cooperative values (Gauthier 1991, p. 27). In numerous passages in *Morals* Gauthier also expresses his confidence in the real-world traction of his contract: in chapter six he says that 'Constrained maximization thus links the idea of morals by agreement to actual moral practice' (Gauthier 1986, p. 168).

[26] '…the force of the social contract is not found simply in its being an agreement. Rather its force lies in its being the nearest approximation to an agreement in a context in which literal agreement is not possible but would be desirable. We cannot literally choose the terms of our interaction, but we can determine what terms we would rationally choose, from an ex ante standpoint that does not privilege the actual course that our interaction has taken' (Gauthier 2013, p. 619).

> …the force of the social contract is not found simply in its being an agreement. Rather its force lies in its being the nearest approximation to an agreement in a context in which literal agreement is not possible but would be desirable. We cannot literally choose the terms of our interaction, but we can determine what terms we would rationally choose, from an ex ante standpoint that does not privilege the actual course that our interaction has taken (Gauthier 2013, p. 619).

For Gauthier any society that purports to be just would need, in order in fact to be just, to pass his contractarian test: '[in] my paper "Political Contractarianism," I sketch some of the features of any society that I argue would in our circumstances not merely pass the contractarian test, but be required in that their absence would fail the test' (Gauthier 2013, p. 620). The particular impulses which Gauthier seems to think would drive real-world persons to look to his contract for consultation help would be their desires to escape the arbitrariness of their day-to-day circumstances and to reflect more thoroughly on their cooperative setups.[27] Thus Gauthier in 'Twenty-Five On' does appear to see his contract as offering consultation guidance, at least in the sense that it provides persons with a better perspective and better cooperative principles than the perspectives and principles to which they are accustomed.[28]

Yet at the same time the relative inability in practice of the Gauthier contract to offer such guidance would make it difficult for real-world agents to employ it in a role as empirically demanding as 'prescription'. In one sense the reason why the Gauthier contract is less useful than it otherwise ought to be if it were prescriptive is that the optimization strategy of 'Twenty-Five On' could not amid many real-world circumstances successfully promote Pareto-dominant participation. The various cooperative circumstances that we encounter in our lives are always to some extent different from each other, and, being different, suggest different cooperation strategies—over and against streamlined Pareto optimization.

Moreover, if Gauthier were to respond that his new model is simply an interesting idealization and thus that it does not offer consultation guidance then he would be rendering himself vulnerable to a second and perhaps more important charge of prescriptive insignificance. I do not believe that Gauthier is willing to bite the prescriptive insignificance bullet because in 'Twenty-Five On' he at one point asserts that 'these realizations do transform, or perhaps bring to the surface, the character of the relationships between persons that are maintained by the existing constraints, so that some of these relationships come to be recognized as coercive' (Gauthier 1991, p. 29). He does appear to intend for his contract to have traction as a prescriptive theory and to be used as a consultation device when real-world persons

[27] This includes especially the setups which are nefarious (Gauthier 1991, p. 29).

[28] Earlier, in *Morals*, Gauthier confirms these sentiments. For instance, in chapter six of *Morals* he says that 'Constrained maximization thus links the idea of morals by agreement to actual moral practice' (Gauthier 1986, p. 168). Elsewhere, he says that 'we may defend actual moral principles by reference to ideal co-operative arrangements, and the closer the principles fit, the stronger the defense' (Gauthier 1991, p. 28).

desire (a) to know more about their cooperative efforts, and also (b) to improve upon such efforts.

To be sure, while it is possible for ideal-level theorists like Gauthier to delineate their decision-making parameters more clearly than they could otherwise do, nevertheless at the same time any idealized theory that purports to have prescriptive value—such as Gauthier appears to want his contract to have—ought at the least to have plausible non-ideal implications. Since few of us are ever going to encounter situations in which our neighbors are universally acting as Pareto optimizers, any implications of Gauthier's idealized beginning would if his theory has practical purchase need doubtless to be credible amid non-ideal circumstances. If and when a particular cooperative model—Gauthier's or otherwise—has dubious real-world implications then presumably, while it might have edification value (viz. like chess or backgammon), its additional value would be limited.

## Conclusion: A Better Way of Interpreting the Gauthier Contract

The upshot is just that I do not think that there is any more interesting real-world consultation help to be found in the revised Gauthier contract than there was in the original in *Morals by Agreement*, and certainly no more than could otherwise be found from more empirically realistic strategies. At first glance it might seem as though second-stage Pareto optimization could produce a process of moral improvement in the American system since that system does seem to need a more widespread choice of second-stage participation in order for its institutions robustly to promote the citizens' welfare. And yet at the same time, upon a closer inspection of the reasoning of the Gauthier contractors, second-stage participation increases are something that the Gauthier contract would arguably be incapable of producing because for many Americans it just is not in any dispositive way rational to be a second-stage participant on a Pareto-optimizing basis—even at the level of dispositions to choose.

The point is also that Gauthier's Pareto optimization strategy is sufficiently otherwise from real-world cooperative forms to be of lesser descriptive or prescriptive consultation value than are other, less abstracted models that are different than the motives of Gauthier's agents and that are less distant from real-world cooperative forms. Since real-world persons are motivated by different and more complicated concerns than the concerns that Gauthier highlights, and since their circumstances are different and more complicated than Gauthier's idealized initial situation, they ought probably to seek their cooperative guidance in reasoning forms that are closer to their existing strategies. In practice this might mean that they should just acknowledge the checkerboard nature of their surroundings and then cooperate in accordance with different principles and for different reasons in the different areas of their lives.

Finally, although Gauthier does at least to a certain extent intend for his contract to have practical traction—and not just to be edifying—I want to stress that I am not here finding fault in the contract's initial setup as such. At the same time, however, my point is that Gauthier is not in the recent article attending to real-world

complexities and that his main goal is no more than to demonstrate the rationality of cooperation, amid a narrow range of idealized circumstances and on a basis that is otherwise than straightforward maximization or a Nash bargaining strategy. As such, in my view maximum use-value would probably be achieved for the revised Gauthier model if it were to be deployed in ways which are empirically modest and attainable, and not for roles that are more demanding than those for which it is equipped.

**Acknowledgments** I am grateful to two anonymous reviewers for their revision suggestions.

## References

Aristotle. 1941, 2001. *Politics, nicomachean ethics*. In *The basic works of Aristotle*, ed. Richard McKeon. New York, NY: Random House Publishers.

Arneson, Richard. 2009. The supposed right to a democratic say. In *Contemporary debates in political philosophy*, ed. Thomas Christiano, and John Christman, 197–212. Oxford: Blackwell.

Barry, Brian. 1995. *Justice as impartiality*. Oxford: Oxford University Press.

Bellah, Robert, et al. 2007. *Habits of the heart: Individualism and commitment in American life*, 2nd ed. Berkeley, CA: University of California Press.

Biel, Anders, Daniel Eek, Tommy Garling, and Mathias Gustafson. 2008. *Promoting cooperation in social dilemmas via fairness norms and group goals*. New York, NY: Springer.

Bogaert, Sandy, Christophe Boone, and Carolyn Declerck. 2008. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology* 47: 453–480.

Braybrooke, David. 1991. Gauthier's foundations for ethics under the test of application. In *Contractarianism and Rational Choice*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Costa, Dora, and Matthew Kahn. 2001. Understanding the decline in social capital, 1952–1998. *National Bureau of Economic Research Working Paper* No. 8295.

Crowder, George. 2002. *Liberalism and value pluralism*. London: Continuum.

Dagger, Richard. 1997. *Civic virtues: Rights, citizenship, and republican liberalism*. Oxford: Oxford University Press.

Declerck, Carolyn H., Christophe Boone, and Griet Emonds. 2013. When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition* 81: 95–117.

Denz, Kimberly. 2012. Young people and public sector careers. Govloop.com. http://www.govloop.com/forum/topics/young-people-and-public-sector-careers. Accessed 3 August 2013.

Dimock, Susan. 1999. Defending non-tuism. *Canadian Journal of Philosophy* 29: 251–274.

Duncan, David. 2012. Why do our best and brightest end up in Silicon Valley and not in D.C.? *The Atlantic*. http://www.theatlantic.com/technology/archive/2012/05/why-do-our-best-and-brightest-end-up-in-silicon-valley-and-not-dc/256767/. Accessed 11 September 2015.

Elshtain, Jean. 1995. *Democracy on trial*. New York: Basic Books.

Galston, William. 2001a. Can patriotism be turned into civic engagement? *The Chronicle of Higher Education* 16 Nov: 16.

Galston, William. 2001b. Political knowledge, political engagement, and civic education. *Annual Review of Political Science* 4: 217–234.

Galston, William. 2004. Civic education and political participation. *Political Science and Politics* 37: 263–266.

Galston, William. 2007. Civic knowledge, civic education, and civic engagement: A summary of recent research. *International Journal of Public Administration* 30: 623–642.

Gans, Herbert. 1991. *Middle american individualism: Political participation and liberal democracy*. Cambridge: Cambridge University Press.

Gauthier, David. 1986. *Morals by agreement*. New York, NY: Oxford University Press.

Gauthier, David. 1993. Between Hobbes and Rawls. In *Rationality, justice and the social contract: Themes from morals by agreement*, ed. David Gauthier, and Robert Sugden, 24–40. Hertfordshire: Harvester Wheatsheaf.
Gauthier, David. 1997. Political contractarianism. *Journal of Political Philosophy* 5: 132–148.
Gauthier, David. 2013. Twenty-five on. *Ethics* 123: 601–624.
Grabb, Edward, et al. 1999. The origins of american individualism: reconsidering the historical evidence. *Canadian Journal of Sociology* 24: 511–533.
Henrich, J., et al. 2005. 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28: 795–855.
Hubin, Donald. 1991. Non-tuism. *Canadian Journal of Philosophy* 21: 441–468.
Knack, Stephen. 2002. Social capital and the quality of government: Evidence from the states. *American Journal of Political Science* 46: 772–785.
Lange, Van, A.M. Paul, Jeff Joireman, Craig D. Parks, and Eric Van Dijk. 2013. The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes* 120: 125–141.
Lapin, Andrew. 2012. Young feds say bureaucracy, inefficiency keeping their peers away. Govexec.com. http://www.govexec.com/management/2012/05/young-feds-say-bureaucracy-inefficiency-keeping-their-peers-away/55618/. Accessed 5 August 2013.
McCracken, Janet, and Bill Shaw. 1995. Social contracts and business ethics. *Business Ethics Quarterly* 5: 297–312.
McClennen, Edward. 2010. Rational choice and moral theory. *Ethical Theory and Moral Practice* 13: 521–540.
Moehler, Michael. 2015. Rational cooperation and the Nash bargaining solution. *Ethical Theory and Moral Practice* 18: 577–594.
Morris, Christopher, and Arthur Ripstein (eds.). 2001. *Practical rationality and preference: essays for David Gauthier*. Cambridge: Cambridge University Press.
Murphy, Ryan, Kurt Ackermann, and Michel Handgraaf. 2011. Measuring social value orientation. *Judgment and Decision Making* 6: 771–781.
Murray, Malcolm. 2007. *The moral wager: Evolution and contract*. New York, NY: Springer.
Putnam, Robert. 1993. *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.
Putnam, Robert. 1995. Tuning in, tuning out: The strange disappearance of social capital in America. *PS. Political Science and Politics* 28: 664–683.
Putnam, Robert. 2000. *Bowling alone*. New York: Simon and Schuster.
Roth, A.E. 1977. Individual rationality and Nash's solution to the bargaining problem. *Mathematics of Operations Research* 2: 64–65.
Schumpeter, Joseph. 2008. *Capitalism, socialism, and democracy*. Harper Perennial Modern Classics.
Southwood, Nicholas. 2010. *Contractualism and the foundations of morality*. Oxford: Oxford University Press.
Rahn, Wendy, and John Transue. 1998. Social trust and value change: The decline of social capital in American Youth, 1976–1995. *Political Psychology* 19: 545–565.
Taylor, Charles. 1985. Atomism. In *Philosophy and the human sciences: Philosophical papers*, vol. 2, 187–210. Cambridge: Cambridge University Press.