30 Years After '*Morals by Agreement*'

MOHAMAD AL-HAKIM *Florida Gulf Coast University*
GARRETT MAC SWEENEY *York University*

We are excited to introduce this special collection of articles written in celebration of the 30th anniversary of David Gauthier's seminal book, *Morals by Agreement* (OUP, 1986).

This special collection of *Dialogue* celebrates the influence Gauthier's work has had on conversations in morality and practical reason, rational choice theory, and the contractarian tradition, and the ways in which his ideas have been refined, corrected, and further supported over the years. Part of the challenge of setting down a collection such as this is that Gauthier's career has been such an illustrious one. So, where shall we start? On which areas of his work shall we focus? To start at the beginning of his career and then to cover it all would lead to a collection far too long for a special edition of a journal. Likewise, Gauthier's influence on a great many different topics and sub-topics within the world of practical philosophy could each produce a special collection in their own right. So we decided to start with what is sometimes thought to be an ideal spot, an anniversary. Anniversaries allow for reflection, allow us to look back at where we have been, ask ourselves what we think worked, what we would have changed, what we think deserves a second look, where we think we want to go, and how we think we need to get there. In such a long, fruitful, and storied career, an anniversary allows us, as celebrants, the opportunity to focus on what we think matters, both in retrospect and moving forward.

The idea for this collection began five years ago during a conference held at York University (Toronto, Canada) that was celebrating Gauthier's book at its 25th anniversary. We were both graduate students at the time, and we were struck by the diverse attendance of a wide range of experts drawn from nearly

***Dialogue* 55 (2016), 593–600.**

doi:10.1017/S0012217316000731

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

every part of the academic spectrum. The conference brought together an impressive group of academics whose work in such fields as economics, evolutionary biology, criminal law, political theory, game theory, and ethics (more broadly) intersected with Gauthier's ideas in one way or another. Some presenters challenged Gauthier's central notion of constrained maximization by way of drawing its limited application to market economics; others defended Gauthier's reliance on the use of such tools as the Prisoner's Dilemma in constructing a theory of morality; others still critiqued or defended the social contract tradition more generally and, in particular, its unique expression set out in *Morals by Agreement.*

According to Gauthier, the ideas articulated in *Morals by Agreement* began nearly 19 years before its publication when, while "fumbling for words in which to express the peculiar relationship between morality and advantage,"[1] he was shown the Prisoner's Dilemma. Although the traditional view interprets the resolution of Prisoner's Dilemmas in favour of straight forward maximization, Gauthier saw the Dilemma as posing a problem concerning practical rationality and cooperation. His project in *Morals by Agreement* was to "provide a justificatory framework for moral behaviour and principles"[2]—a sort of justification grounded in rational choice. Gauthier's central claim is that, in situations involving interactions with others, "an individual chooses rationally only in so far as he constrains his pursuit of his own interest or advantage to conform to principles expressing the impartiality characteristic of morality."[3] In order to meet the challenge he saw associated with the Prisoner's Dilemma, Gauthier addressed three interrelated core problems. The first concerned the need to formulate the principles of rational cooperation (i.e., a constrained maximization). The second core problem concerned the need for demonstrating the rationality with complying with the selected principle. Finally, the third problem was to "determine the appropriate initial position from which co-operation proceeds."[4] Providing a response to the three core problems motivated Gauthier to develop several central elements of his brand of contractarianism. One such element is the idea that what economists call a perfectly free and competitive market is a morally free zone. Others include the core principles for determining how the benefits of cooperation should be shared (i.e., minimax relative concession and maximin relative benefit), a developed conception of constrained maximization, and finally, a (Lockean) proviso for establishing the initial conditions from which agreement can be fairly negotiated.[5] Part of what makes Gauthier's approach unique is that it fundamentally breaks

[1] Gauthier, 1986, p. v.

[2] Gauthier, 1986, p. 2.

[3] Gauthier, 1986, p. 4.

[4] Gauthier, 1986, p. v.

[5] Gauthier, 1986, pp. 13-17.

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

from the dominant social contract view developed by John Rawls. Although Gauthier admits that Rawls's idea—namely, "that principles of justice are the object of a rational choice"[6]—is incorporated into his own theory, there remain important differences. In particular, Gauthier insists that his theory claims to "generate morality as a set of rational principles for choice"; he takes his task to be showing "why an individual, reasoning from non-moral premises, would accept the constraints of morality on his choice."[7] This theory contrasts with Rawls's project, which sets out to determine the shared principles of justice that would be agreed upon from a fair initial bargaining position behind the veil of ignorance and whose two principles of justice act as constraints on social institutions. What we instead get in *Morals by Agreement* is nothing short of a substantive and fully worked out theory of rationality and its relationship to morality.

Since first appearing 30 years ago, *Morals by Agreement* has heavily influenced the way philosophers, economists, political theorists, and evolutionary biologists think about the basis of social cooperation and its related parts. Over the last three decades, important advances have been made in the fields of evolutionary psychology, moral theory, economics, game theory, rational choice theory, the study of practical reason, the study of the moral conditions for criminal law, business ethics, and political science. Many working in these areas have drawn on Gauthier's insights and his framework in advancing their own analyses, while others have challenged some of his central assumptions about human nature, moral motivation, and the role of constrained maximization in explaining cooperative behaviour. The influence of *Morals by Agreement* has endured, and continues to be explored by theorists in a variety of areas and disciplines.

Many theorists ask whether rational agents actually go about making decisions in the way Gauthier assumes. What limits are there to adopting a Hobbesian view of rational agency? What implications does Gauthier's work have on schemes of distributive justice? Evolutionary biology? What other factors might explain moral constraints on the behaviour of individuals? Do coordination problems get resolved in the way Gauthier suggests, or are there other explanations for agent behaviour not accounted for by Gauthier? What might such factors look like and what do they tell us about rational agency? What is the appropriate role for the Prisoner's Dilemma in moral theorizing? Can all of this solve the compliance problem? These questions form the foundation of this collection. The featured authors have all, in one way or another, engaged with either the questions or the answers presented by Gauthier over the last 30 years.

The following collection consists of eight new articles, including a long-awaited description by Gauthier of what a contractarian society of the type his

[6] Gauthier, 1986, p. 5.

[7] Gauthier, 1986, p. 5.

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

theory would endorse might look like in the real world. The papers were derived through a mixed process: submissions for the conference at York University were peer reviewed, and we subsequently issued a wider call for submissions on the theme of '30 Years after *Morals by Agreement.*' We wish to emphasize that the following collection is not intended to be a restricted or parochial engagement with Gauthier's central text. Rather, *Morals by Agreement* serves as the initial point for each article, with each article then developing a critique or expansion of Gauthier's ideas and rational choice contractarianism in general.

In these introductory comments, we have divided the articles into three general themes: critics, defenders, and general commentary on Gauthier's ideas. The critics (Andreou, Mullins, and Viminitz) all offer different points of criticism of Gauthier's work. In her article, "Figuring Out How to Proceed with Evaluation After Figuring Out What Matters," Chrisoula Andreou revisits Gauthier's discussion of agency, action, and motivation. Gauthier generally holds that the rationality of an action is not to be determined directly by the outcomes it produces but instead through an evaluation of the deliberative procedure the agent employed in reaching the decision. In other words, deliberative procedures are the primary objects of evaluation for Gauthier. Andreou questions the viability of this move by suggesting that "even the most direct evaluation of intentional actions involves the evaluation of different ways of deliberating about what to do." This opens the door for a more holistic approach to the evaluation of rational deliberation, one that goes beyond the orthodoxy's focus on choice at a given 'time-slice.'

One area of study heavily influenced by and engaged with Gauthier's work is economics and market-contractarianism (i.e., the view that market decisions are purely motivated by instrumental considerations). In his article, "Gauthier, Equilibrium, and the Emergence of Morality," Brett Mullins revisits two principles centrally tied to the emergence of market morality, which he calls 'Strategic Emergence' and 'Market Emergence.' Gauthier's theory is intended to rationally motivate a constrained form of morality in the public sphere. These two principles speak to the ways that market morality emerges. The former suggests that morality emerges just in case strategic equilibrium is not optimal, while the latter, Market Emergence, emerges just in case market failures obtain. Mullins questions the consistency of these two principles and the implications for dropping one (or both) from Gauthier's market model. Mullins suggests that Gauthier's theory resists either option and as a result fails, in this respect, to rationally motivate morality.

Paul Viminitz's thought-provoking paper, "Getting the Baseline Right—or—Why I'm Right and Everyone Else is Wrong, in each of the Two Senses of 'Why,'" offers a broad general critical approach to some of the central concepts developed by Gauthier since the publication of *Morals by Agreement.* While Andreou and Mullins both take up specific critiques to elements in Gauthier's theory, they do so with an eye to expanding his ideas and finding

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

formulations that better fulfill his theoretical objectives. Viminitz, by contrast, challenges a specific element in Gauthier's theory, namely, the Lockean proviso. The proviso governs the conduct of would-be cooperators prior to their coming together to form the social contract. It prohibits individuals from benefitting themselves (or others) *by* worsening the situation of others, and then using their improved position to demand a greater share of the cooperative surplus (the goods created by cooperating over and above what the individuals could have created acting alone) than is afforded to the others. The proviso specifies a pre-moral state, the baseline, from which selecting the terms of social cooperation can occur. Viminitz argues that the baseline of social cooperation has been fundamentally misunderstood and misplaced. Moreover, Viminitz's contention is not restricted to Gauthier but also extends to other important thinkers in the field, including Jan Narveson (who is also featured in this collection). His conclusion is not that Gauthier and Narveson are actually wrong, but rather that they are mistaken about the reason they are right. Viminitz's paper serves both as a critique and as an attempt to tease out some of the important differences among rational choice contractarians and himself. It also offers a thoughtful and entertaining (but serious) engagement with Gauthier.

The next set of articles can be generally grouped under the heading of 'defenders.' Like their counterparts, the defenders (Cohen, Kuhn, and Narveson) all offer a specific or general defence of Gauthier's central ideas and the social contract tradition more broadly. In "Contractarianism and Moral Standing Inegalitarianism," Andrew Cohen takes up the important question of the inclusiveness of contractarianism in terms of its ability to justify equal moral standing of persons as well as to provide grounds for the moral standing for many non-human animals. The latter is easier to comprehend, since Gauthier's theory is premised on rationality as the source of constrained morality. This seems to preclude the possibility of including non-human animals as well as other potential human sub-groups, such as future generations or current members incapable of social cooperation from the terms of the social contract. Does Gauthier's model support an inegalitarian standing of individuals based on the criterion of rationality? Cohen teases out the relationship between contractarianism, liberalism, and egalitarian concerns. He concludes with the suggestion that, while contractarianism may permit for some entities to have more moral standing than others, this does not license the sort of oppression liberal egalitarians often rightly fear.

One of the central running examples utilized by Gauthier for his moral theorizing is the Prisoner's Dilemma. Despite its centrality in Gauthier's writing, critics have questioned the appropriateness of using it to think about morality. Critics, such as Kenneth Binmore, for example, suggest that it is "just plain wrong to claim that the Prisoners' Dilemma embodies the essence of the game of human cooperation." Rather, it instead "represents a situation in which the dice are as loaded against the emergence of cooperation as they could possibly be," thereby conflicting with evolutionary accounts of cooperative schemes

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

and behaviours. The seriousness of this (and other similar) criticism cannot be understated as it attempts to undermine one of the most substantive elements of Gauthier's theory. Steven Kuhn's paper, "Gauthier and the Prisoner's Dilemma," traces the evolving role of the Prisoner's Dilemma in Gauthier's work from a 'model' in one of his earliest publications[8] to the more specialized role of helping to situate the difficulties with the view that moral action is individually rational in *Morals by Agreement*. More specifically, Kuhn's project is twofold. First, he defends Gauthier's use of the Prisoner's Dilemma as a model for moral theorizing, and second, he presents a sketch of a more developed descriptive and normative account of rational choice contractarianism that derives its bases from Gauthier but also develops aspects of Kuhn's own work that were previously undeveloped.

Jan Narveson's contribution, "Social Contract: The Only Game in Town," offers an outright defence of the social contract tradition, and, more specifically, the core idea of 'rationality,' as developed in the tradition and by Gauthier in particular. Narveson, like Viminitz, shares his thoughts on the value of social contract theory, going so far as to declare that the approach is indeed, as Gauthier once remarked, 'the only game in town.' Narveson commits much of his article to vindicating two main elements of Gauthier's book, namely the notion of constrained maximization and the reliance on the Lockean proviso. His article offers a rich historical and philosophical analysis of the rational choice contractarianism of Gauthier, the central problems it raises for issues of distributive justice, and an overall defence of the tradition as the most viable way to think about and construct a system of shared morality. It is a provocative paper that is sure to motivate readers to think through the potential implications of maintaining an ethics based on mutual self-interest and one which grounds moral obligations in their consistency with rationality.

In addition to the above contributions, this collection also includes a paper that draws out some of the connections between Gauthier's work and other specialized work in game theory and cooperative schemes more generally. To this end, Robert Sugden's contribution, "On David Gauthier's Theories of Coordination and Cooperation," offers a refreshing analysis of the core features of Gauthier's book often less discussed by philosophers though widely taken up by economists, mathematicians, psychologists, and game theorists. Sugden focuses his contribution on a less well-known paper by Gauthier simply titled "Coordination."[9] In this paper, Gauthier takes up two well-known games, the *pure coordination game* (first described by Thomas Schelling) and *Hi-Lo game* (whose paradoxical features were outlined by David Hodgson). Part of Gauthier's purpose was to distinguish his approach from those of Schelling and Hodgson and to highlight an important distinction between 'coordination' (as exemplified by

[8] Gauthier, 1967.

[9] Gauthier, 1975.

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

pure coordination and Hi-Lo games) and 'cooperation' (as exemplified by the Prisoner's Dilemma). Sugden revisits Gauthier, Schelling, and Hodgson in an attempt to further situate the tension between Gauthier's proposed view of social cooperation and those of Schelling and Hodgson. Rather than merely explain the tension, however, Sugden draws out a significant parallel between Hodgson's claims that "rational decision-making in games can be construed in terms of the players jointly choosing the combination of strategies that is best for them collectively, rather than … each choosing the strategy that he individually judges to be best," and Gauthier's notion of constrained maximization. Sugden argues that Gauthier's fundamental insight about game theory would be better served by erasing the cooperation/coordination distinction that he relies upon and instead applying the notion of constrained maximization to both domains. The upshot of this is that it helps resolve the tension between Gauthier's view and Hodgson's. Moreover, Sugden suggests that doing so "allows one to see Gauthier's conception of rationality as a distinctive and attractive form of 'team-reasoning.'"

Lastly, we are fortunate enough to have the opportunity to include a new article by David Gauthier, written specifically for this special collection. We asked Gauthier if he would be generous enough to produce a reflective piece that revisits many of the main ideas developed in *Morals by Agreement* as well as to offer, if possible, a sketch of where the theory might be headed next. We are excited to share his special contribution.

In the Preface to *Morals by Agreement*, Gauthier opens with the following personal words:

> The present enquiry began on a November afternoon in Los Angeles when, fumbling for words in which to express the peculiar relationship between morality and advantage, I was shown the Prisoner's Dilemma. Almost nineteen years later, I reflect on the course of a voyage that is not, and cannot be, completed, but that finds a temporary harbor in this book.[10]

The culmination of that 19-year voyage is not noted as an ending, however, but instead as a beginning. As Gauthier writes at the close of the Preface,

> And so I come to an end, aware that it is also a beginning, for I shall surely find myself embarked again on the quest to understand how morality and rationality are related."[11]

In many ways, Gauthier's influential ideas began with a sketch. This sketch has since become a theory in its own right and a strong contender within the contractarian tradition. However, it is difficult to know when a sketch is fully,

[10] Gauthier, 1986, p. v.

[11] Gauthier, 1986, p. vi.

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press

if ever, complete. For what can be made can often be bettered, and what is at one time not considered or foreseen can come to alter the very grounds from which a theory springs. Gauthier's new contribution is a kind and humbling reminder that the work of a theorist, even one as accomplished as Gauthier, is never complete. What Gauthier leaves us with in this collection is a new sketch of ideas, much like the sketch of ideas he began with 19 years before the publication of *Morals by Agreement*. We hope that Gauthier's reflection on his work, including his new sketch of "A Society of Individuals," will provide the motivation for theorists working in the field to once again embark "on the quest to understand how morality and rationality are related."

We think it is best to leave the final words of this introduction to Gauthier's own description of his paper, one that we think will provide fuel for the present and next generation of rational choice contractarians.

> In "A Society of Individuals," I sketch a society that has no good of its own, no social end, but exists to enable each individual member better to pursue his own good, facilitating cooperation, and resolving the basic Interaction Problem (exemplified by the Prisoner's Dilemma): that utility-maximization and Pareto-optimization are sometimes incompatible. The orthodox defend the rationality of maximization; I defend Pareto-optimization. I argue that if (*per impossible*) we could determine the features of our society by prior agreement we would agree to a Society of Individuals, and that we would agree *ex ante* to some social practice or institution is the best possible justification of it holding for us.
>
> I then sketch some of the main features of the Society. In doing this I assume that members of the Society are not all adherents of contractarianism, but may hold any of a number of reasonable views, which the Society must seek to accommodate. I consider how several alleged rights, such as a right to resources, fare in the Society. And I conclude with the idea that contractarianism, in arguing that each adult member of society enjoys equal citizenship, must afford each the right to participate in choosing and dismissing governments. We may then think the emergence of a Society of Individuals is democracy's fulfillment.

References

Gauthier, David

1967 "Morality and Advantage," *Philosophical Review* 76 (4): 460–475.

Gauthier, David

1975 "Coordination," *Dialogue* 14 (2): 195–221.

Gauthier, David

1986 *Morals by Agreement*, New York: Oxford University Press.

https://doi.org/10.1017/S0012217316000731 Published online by Cambridge University Press