

Creative Commons Attribution 4.0 International (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Economic Modeling in Rawls: The Original Position

DAVID C. COKER

University of Maryland, Baltimore County

Abstract: Critics of Rawls's *A Theory of Justice* frequently envision his original position as containing a human consciousness. Thus, the restrictions Rawls introduces for this 'individual'—the lack of particular circumstantial and personal information—is considered a potential problem. The very ways in which Rawls circumscribes the knowledge available in this position is thought to compromise the personhood of the individual there, and hence as well the conclusions reached (that is, Rawls's two principles). This paper will argue that, on the contrary, the lack of full personhood is a critical part of Rawls's modeling strategy, and that Rawls borrowed this particular sense of modeling from economics. It is well known that Rawls worked to verse himself in economic theory, and it is difficult to overlook its use in *Theory*. It will be argued that it is through parallels with economic reasoning that Rawls's original position model can be most fruitfully understood.

Keywords: Rawls, original position, economics, model, Mill, Kant, rationality, *homo economicus*

JEL Classification: B20, B31, B40, H30

I. INTRODUCTION

Who are the individuals in Rawls's 'original position'? A reader's answer to this question frequently determines a judgment on Rawls's entire system. In the original position, Rawls posits that individuals are stripped of much of their individuation, to achieve a certain end: "One excludes the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices" (*A Theory of Justice*, 19).¹

¹ This and all subsequent references to *A Theory of Justice*, abbreviated as '*TJ*', will be to the original edition (Rawls 1971).

AUTHOR'S NOTE: I want to thank David Levy for the initial inspiration to work on Rawls, as well as direction towards particular sources. I must also thank Thomas Scanlon for helpful criticism, and encouragement. And the referees for this journal deserve particular mention for comments which dramatically added to the quality of the presentation.

The problem is: after one is deprived of knowing one's individual history, awareness of social position, wealth, idea of possible prospects, race, and sex, is there sufficient *self* left to make reasonable choices about social and moral guidelines? On the face of it, it is perplexing. Are human beings behind the veil of ignorance real enough that we feel persuaded that their decisions would be our own?

This paper will argue that this line of questioning is fundamentally misconceived. However, there are two separate contexts to which such a claim may apply: the Rawls of *A Theory of Justice* (and before),² and the Rawls of *Political Liberalism*.³ Rawls clearly had misgivings about the level of reliance on the rationality claim, and many of the critical building blocks utilized from economics, by the time of *PL*.⁴ Such refinements of the theory could be considered advances, or not.⁵ The focus here will be on Rawls previous to *PL*, primarily the Rawls of *TJ*. Even within *TJ*, however, there exists a troubling variety of terms of description. Rawls discusses the original position in two largely contradictory ways. One side of his presentation invites our participation, our projection of ourselves behind the veil—a kind of humanizing of his system. But another type of description—which Rawls takes some pains to emphasize is primary—is that the entire original position, including the ‘individuals’ situated there, is engineered. It is, essentially, a machine designed to generate a certain sort of output. This mechanistic sense of how the original position functions has proven an enormous stumbling block for critics. One dimension of the problem for critics is that Rawls developed this mechanistic formulation—which he had from his undergraduate days—through his close study of economics. This paper will attempt to make the argument that the original position is mechanical (Rawls’s in-

² It is likely that this cut-off is too severe. The lectures in *Justice as Fairness: A Restatement*, which are from the 1980s, do appear very much in sync with *A Theory of Justice*.

³ This and all subsequent references to *Political Liberalism*, abbreviated as ‘*PL*’, will be to the expanded edition (Rawls 2005).

⁴ “I use the distinction between the two parts of the original position which correspond to the reasonable and the rational as a vivid way to state the ideas that this position models the *full* conception of the person. I hope that this will prevent several misinterpretations of this position, for example, that it is intended to be morally neutral, or that it models only the notion of rationality, and therefore that justice as fairness attempts to select principles of justice purely on the basis of conception of rational choice as understood in economics or decision theory. For a Kantian view, such an attempt is out of the question and is incompatible with its conception of the person” (*PL*, 306n21).

⁵ For instance, T.M. Scanlon favors the Rawls of *TJ* (personal correspondence). For a compelling overview favoring the later Rawls, see Audard (2007).

tention), and examine how the parallels with economic theory make this more easily seen.

One predisposition that places us in a position to see Rawls's idea more clearly is to envision the original position as a model. Models in economics have an occasionally shocking level of abstraction. The pushback against modeling in economics is similar to what we see in studies of Rawls; the individual (*homo economicus*) fails to tally with anything like our introspective view of ourselves. The modeled person lacks dimension. He or she also is assumed to act in isolation—decisions are assumed independent from those of others. How can the model be informative—how can it pertain to science—while misrepresenting human nature so completely? Such questions should be familiar to readers of and commentators on Rawls. His modeled 'person' is also stripped of critical dimensions and decides in a state of isolation. Yet these harsh caricatures—models—in economics, though being adjusted at the margin in various ways, have survived. Their survival is tied to what they offer the theorist. Economies are vast and complicated entities, and simplifications are required to add coherence to even a single element in the process. The complications in social ordering and morality—the issues Rawls faces—are also hugely complex. Rawls appreciates how models work in economics, and it is our perception of that appreciation that enables us to see his original position in the terms he intends.

II. USE OF MODELS

The approach of this paper—the use of neighboring disciplines to add dimension and insight to one another—is a technique that Rawls himself utilized. This is, of course, part of a larger argument about the importance of economic thinking in Rawls generally. Rawls explicitly invokes economics throughout *A Theory of Justice* and *Justice as Fairness: A Restatement* (2001). And he also argues in his lectures for the overlap of economics and political philosophy being a fruitful one. For instance, in his chapter on Hume's "Of the Original Contract", Rawls discusses the advantages of this dual attack:

Still, since 1900 the tradition [of utilitarian analysis] has divided into two more or less mutually-ignoring groups, the economists and the philosophers, to the reciprocal disadvantages of both; at least in so far as economists concern themselves with political economy and so-called welfare economics, and philosophers with moral and political philosophy. (Rawls 2007, 162–163)

It is important to realize that one of Rawls's early inspirations—Mill's utilitarianism—predates the schism mentioned in the quote. Mill was, we know, both a social theorist, facing the questions which interested Rawls, as well as the most famous economist of his era. And Mill had his own definition of 'economic man' (*homo economicus*). That definition centered on seeing individuals as wealth-accumulators. "Mill thus constructs a *homo economicus* but does so fully aware that his artifice is an ideal type which rarely has its exact counterpart in the world of reality" (Spiegel 1983, 380). Mill explains this in more depth:

No mathematician ever thought that his definition of a line corresponded to an actual line. As little did any political economist ever imagine that real men had no object of desire but wealth, or none which would not give way to the slightest motive of a pecuniary kind. But they were justified in assuming this, for the purposes of their argument; because they had to do only with those parts of human conduct which have pecuniary advantage for their direct and principal object; and because, as no two individual cases are exactly alike, no *general* maxims could ever be laid down unless *some* of the circumstances of the particular case were left out of consideration. (Mill [1967] 2006, 327)

Rawls, looking for a path into his problem, was certainly aware of these simplifications and abstractions Mill describes. They would provide an example of, and demonstrate the potential benefits of, this sort of modeling of the individual. And Rawls's explorations of economics would bring him into contact with even more severe economic modeling. By the time one reaches the neoclassical view of *homo economicus* present in Knight (as opposed to the classical one in Mill), one finds that the *purely* economic actor is not human at all, but in his mechanical responses is more akin to a slot machine!⁶

In his economic reading—and Rawls footnotes over forty economists in *TJ* alone—Rawls would have come across a great deal of such model-making. Mary Morgan summarizes this tendency in 19th century economics:

But these were all model men compared to the rich descriptive portrait we find in other works of social science. Each model man was made to reduce the complexity of dealing with all human feelings

⁶ This is Knight's actual phrase. For a more in-depth look at Knight's influence on Rawls, see Coker (2021).

and emotions and actions that flow from them and, at the same time to focus the attention on the explicitly economic aspects of man's behaviour. This sequence of model men was the nineteenth-century economists' answer to the problem of dealing with human behaviours in a scientific way. (Morgan 2012, 164–165)

This tendency towards abstraction would only become more extreme in the contemporary economists with whom Rawls was also familiar. And Rawls himself, as we will see, will take great liberties with the notion of 'rich descriptive portraits' that some readers will expect to animate his notion of individuals behind the veil.

Before moving to the particulars of the argument, it might be useful to briefly touch on Rawls's use of economics as a reference point. Such references, no matter Rawls's reliance on them, tend to be excluded from critics' summaries of his arguments. For instance, when Rawls considers just rewards for labor, his thinking and terms of analysis are economic:

It is easy to see, however, that this is not the case. The marginal product of labor depends upon supply and demand. What an individual contributes by his work varies with the demand of firms for his skills, and this in turn varies with the demand for the products of firms. An individual's contribution is also affected by how many offer similar talents. There is no presumption, then, that following the precept of contribution leads to a just outcome unless the underlying market forces, and the availability of opportunities which they reflect, are appropriately regulated. (*TJ*, 308)

More startlingly, the reference point in discussing politics is again economic:

The ideal procedure is further clarified by noting that it stands in contrast to the ideal market process. Thus, granting that the classical assumptions for perfect competition hold, and that there are no external economies or diseconomies, and the like, an efficient economic configuration results. The ideal market is a perfect procedure with respect to efficiency. (*TJ*, 359)

By the next page we will find out that politics, of course, is *not* a perfect procedure. Discussing the original position, he once more centers it on economics:

I now turn to matters of detail. Note first the similarity between arguments from the original position and arguments in economics and social theory. The elementary theory of the consumer (the household) contains many examples of the latter. In each case we have rational persons (or agents) making decisions, or arriving at agreements, subject to certain conditions. (Rawls 2001, 81)

The reference is to constrained maximization, a basic technique in economics. He will go on to discuss the ability to predict the economic actor's choices. He soon turns to the need to avoid depending on external "psychological hypotheses or social conditions not already included in the description of the original position" (82). He elucidates the idea by discussing an economic example. The passage continues:

Consider the proposition in economics that the agent for the household buys the commodity-bundle indicated by the (unique) point in commodity-space at which the budget line is tangent to the (highest) indifference curve touching that line. This proposition follows deductively from the premises of demand theory. The necessary psychology is already included in those premises. Ideally we want the same to be true of the argument from the original position. (82–83)

The list of examples could be extended considerably, but these should serve to establish two points. The first is simply Rawls's familiarity with, and use as a reference point of, economics. His recourse to economics as a frame of reference, and as a technique of reasoning, is frequent and continuous. Rawls, to a surprising extent, *sees into* his problems through an economic lens. The second is how economics, by fashioning the terms of analysis—by its assumptions—is able to point to outcomes. It is this sense of modeling that helps us see how Rawls understands his original position. The restrictions on decision-making in the original position are extreme. These restrictions have caused many of Rawls's commentators to balk, insisting that violence has been done to critical components of human nature. But in the context of economic modeling that Rawls saw so plainly, restrictions/assumptions dispense with fidelity to reality as a standard, and instead are a closed system with the outcomes they produce. Rawls insists that he constructs the original position exactly to achieve/predict a particular outcome—his principles. Yet the full implications of his model understood purely in those terms remains a challenge to conceptualize.

III. 'INDIVIDUALS' IN THE ORIGINAL POSITION ARE NOT INDIVIDUALS

Rawls is specific in describing his theoretical procedure as modeling. Answering the charge that the hypothetical agreements in the original position would have no 'significance', he argues:

In reply, the significance of the original position lies in the fact that it is a device of representation or, alternatively, a thought-experiment for the purpose of public- and self-clarification. We are to think of it as modeling two things. (Rawls 2001, 17)

He then details those two things: fair conditions of agreement and "acceptable restrictions on the reasons on the basis of which the parties, situated in fair conditions, may properly put forward certain principles of political justice and reject others" (17).⁷ Even more explicitly:

In using the conception of citizens as free and equal persons we abstract from various features of the social world and idealize in certain ways. This brings out one role of abstract conceptions: they are used to gain a clear and uncluttered view of a question seen as fundamental. (8)

Or, keying off the parallels with economics (partially quoted earlier):

The proposition follows deductively from the premises of demand theory. The necessary psychology is already included in those premises. Ideally we want the same to be true of the argument from the original position: we include the necessary psychology in the description of the parties as rational representatives [...]. As such, the parties are artificial persons, merely inhabitants of our device of representation: they are characters who have a part in the play of our thought-experiment. (83)

Rawls is here sculpting what he calls individuals merely to produce a certain result. This sort of paring away of the unhelpful, the intractable, and the irrelevant characterizes much of economic thinking. Here is Hal Varian's introduction in his *Microeconomic Analysis*, where he discusses equilibrium:

⁷ Here is a perfect example of Rawls's ambiguity of presentation. In *TJ* as well, he states how in the original position "principles of justice should be chosen under certain conditions". But later in that same paragraph, he says, "the ideal outcome would be that these conditions determine a unique set of principles" (*TJ*, 18). In this later refinement, it is no longer individuals choosing, but conditions determining.

The second analytic technique that we will use in our study of microeconomic behavior involves the study of *equilibrium*. At its broadest level, equilibrium analysis can be viewed as the analysis of what happens to an economic system when all of the unit's behavior is *compatible*. Thus we will typically not be concerned with the analysis of an economic system when some firms or consumers find their actions thwarted.

The focus on equilibrium analysis is not due to the belief that equilibrium is necessarily more important than disequilibrium, but rather that the analysis of behavior in disequilibrium is substantially more difficult. (Varian 1984, 1)

Here is a classic economic assumption, seeking an answer or conclusion through a radical simplification of circumstances, and letting those simplifications determine theory. It could be, and has been, argued that excluding situations in disequilibrium essentially excludes economic situations altogether. There *is* pushback in economics against how relevant equilibrium analysis is for real-world applications. But economic theory is held together by a web of such models, mathematical and otherwise. Morgan indicates that modeling of a technical sort began to dominate economics in the latter half of the twentieth century. But in the more general sense in which we are considering it here, it has been central from at least the time of Adam Smith. It represents a group of keys that economists try in every lock. And its ascendancy as a self-consciously used methodology parallels the time of Rawls's early exposure to it. By then it "had become *the* accepted mode of reasoning in economics in the sense that it became 'the right way to reason [...] what it is to reason rightly'" (Morgan 2012, 13–14). Economists in the 1950s and '60s were beginning to address questions that interested Rawls, bringing with them a mode of thinking closely linked to their training in this intensive modeling.

This has all been somewhat hypothetical and general so far. How would it look in terms of a specific critique of Rawls's position? In his *Liberalism and the Limits of Justice*,⁸ Sandel (1982) argues that the results derived from the original position have a certain reflexive character. The resulting final product is one "of dual dimensions, and in this it's the key to our account", Sandel argues, and continues:

⁸ Focusing on this particular issue in Sandel should not imply there are not interesting insights elsewhere in his study!

For what issues at one end in a theory of justice must issue at the other in a theory of the person, or more precisely, a theory of the moral subject. Looking from one direction through the lens of the original position we see the two principles of justice; looking from the other direction we see a reflection of ourselves. If the method of relative equilibrium operates with the symmetry Rawls ascribes to it, the original position must produce not only a moral theory but also a philosophical anthropology. (Sandel 1982, 48)

The passage in Rawls that is mentioned as the springboard for this view is the section in which Rawls discusses going in and out of the original position, as a kind of check on what we are thinking there (*TJ*, 20). This movement is at least partly to evaluate the reasonableness and strength (or weakness) of the principles generated. Rawls argues that we wish these principles to be as weak as possible but still answer to our considered moral judgments. They must also be acceptable to others. But this process is to modify the original position itself; if we cannot reach ‘significant’ judgments, then we should adjust the conditions under which we deliberate.

Sandel’s assessment of what should be happening here requires a jump in reasoning. He takes this experimental back-and-forth Rawls describes as implying that we are judging the evaluative self as well. It is not just the deliberative procedure that is in play, it is the sense of ourselves as actual individuals behind the veil. There is, as he puts it, an anthropology involved.

We must be prepared to live with the vision contained in the original position, mutual disinterest and all, prepared to live with it in the sense of accepting its description as an accurate reflection of human moral circumstance, consistent with our understanding of ourselves. (Sandel 1982, 48–49)

Rawls, though, does *not* intend this ‘vision’ to be ‘an accurate reflection of human moral circumstance, consistent with our understanding of ourselves’, in anything like the sense that Sandel wishes to understand it here. Sandel will go on to argue that surely we cannot be content to have this “individual” so cut off from society and its effects; “it rules out the possibility of a public life in which, for good or ill, the identity as well as the interests of the participants could be at stake” (62). The reason we are not compelled by such an argument is that it has disconnected from Rawls’s text altogether. Rawls might argue that such things have already

had their effect, and have been incorporated in our deepest moral intuitions. But for our purposes Sandel's error is to take the *model* for a depiction of an actual individual in some critical way. And as often as Rawls uses the word model to describe his constructs, Sandel (and many other critics) tellingly fail to follow suit. Models are *functional*, and they will be unrealistic to the degree necessary to achieve that functionality. This should instill a degree of caution in evaluating them. And standards of 'richness' and/or verisimilitude would likely be particularly fraught starting points. Rawls is elusive here; *he utilizes models in a manner economists take as standard, but which philosophers frequently attempt to understand on other grounds*. The result, in criticism such as that offered by Sandel, is that the argument misses its subject. (Economists, unfamiliar with many philosophical premises, also often transpose Rawls's arguments along lines more compatible with *their* training, with similar results.)

IV. CAN THE ORIGINAL POSITION BE SIMULATED IN EVERYDAY LIFE?

Much of the discussion so far has been on the portrait of the 'individual' behind the veil, and to what standards of realism or completeness that portrait should be held. But modeling is two-sided: it sacrifices the verisimilitude of the agent, to achieve a predictable outcome. The outcome, in that sense, is the driver; the assumptions, and the agent, are molded to whatever degree necessary to achieve that outcome. Rawls is direct in indicating that this is how he sees the original position—as a situation designed to output certain principles. Yet Rawls places the original position in a variety of contexts. While economics is clearly a central one, the inclination to place it in relation to Kant is present in *TJ*, and continues through the various refinements he introduces later. How does, or does, the mechanistic model conform to a Kantian connection?⁹ Rawls suggests that the noumenal self from Kant finds its corresponding entity in the 'individual' behind the veil. This claim of linkage puts stress on the system. The argument in *TJ* is that 'acting autonomously' in Rawls's theory plays out in the nature of the original position.¹⁰ The Kantian (noumenal) self would not be influenced or controlled by circumstance in choosing moral boundaries and priorities; the original position suitably isolates the decision from these elements. Yet this way of arguing

⁹ Argued in (*TJ*, 251–257).

¹⁰ Rawls also links the veil of ignorance to Kant, saying it is "implicit" in Kant's ethics (*TJ*, 140–141; §40).

might seem to place an actual self behind the veil, something that Rawls argues explicitly against elsewhere. One way to understand this difficulty is to recall Rawls's use of—in his economic parallel—of a psychology being ‘built in’. “The proposition follows deductively from the premises of demand theory. The necessary psychology is already included in those premises” (Rawls 2001, 83). In the original position, Rawls implies the similar inclusion of a ‘necessary psychology’. Just as the ‘individual’ in demand theory in economics is pared down to her dimension as ‘consumer’, Rawls's agent is likewise pared to only those elements required for ‘moral choice’. That this agent, as in economics, is not actually a person implies a certain interpretation of Kant's noumenal self. In “Kantian Constructivism in Moral Theory”, Rawls ([1980] 1999) does discuss ‘moral persons’ behind the veil. But the terms appear unchanged from *TJ*; Rawls claims that “I am concerned with the parties in the original position only as rationally autonomous agents of construction” (308). These are not actual persons: “The rational autonomy of the parties in the original position contrasts with the full autonomy of citizens in society” (308). Later he reiterates that the parties “are rational agents of construction” (315). We must leave it there.¹¹ We fall back on “at the basis of the theory, one tries to assume as little as possible” (*TJ*, 129). But certainly Rawls's engagement with Kant was involved and ongoing.

Rawls's quotation above—concerning the ‘necessary psychology’ being built-in—does require a clarification of how Rawls intends ‘prediction’ in his theory to be understood. The classic paper on prediction in economics is probably Friedman ([1953] 1966), where Friedman famously separates reasonable assumptions from accurate predictions, insisting that only the accuracy of the prediction, not the reasonableness of the assumptions, are to guide us in selecting a model. Rawls, in a book (*TJ*) awash with citations from economics, glaringly fails to cite this paper. This should reinforce our sense that Rawls's use of prediction is distinct from Friedman's, and from econometric usage generally. Friedman's cavalier attitude toward assumptions doesn't match up well with Rawls's carefully deliberated original position. Rawls's mention of demand theory, perfect competition, and of tangencies of curves speaks of

¹¹ Taylor (2014) argues similarly from later Rawls ([1980] 1999). Taylor discusses the second moral power of rationality, defined by Rawls as the “capacity to form, to revise, and rationally to pursue a conception of the good” ([1980] 1999, 312): “Thus, Rawls's second moral power of rationality which unites deliberative rationality with creative self-authorship, is simply a variation on the contemporary concept of *personal* autonomy. [...] This self-authorship links rationality (so understood) to Kantian *moral* autonomy, which Rawls endorses in *Theory*, §40” (Taylor 2014, 154–155).

a different type of parallel. In those instances, the level of abstraction is severe. There is no ‘self’ for the individuals in those scenarios, there is only ‘outcome’. This is clearly what Rawls aspires to for his theory. As an example, one might imagine being a consumer facing a certain market price. We enter imaginatively into the role of the consumer, yet the construction exists outside this entering, and does not depend in any way upon it. This is the proper way to understand the original position. We “can by deliberately following the constraints it expresses simulate the reflections of the parties” (*TJ*, 120). But despite our negative or positive reaction during this simulation, the structure of the original position and the principles it generates remain unchanged.

But there *is* adjustment, or flux, present in how the individual is assumed to arrive at reflective equilibrium. The process there is matching and modifying moral dispositions; we search for weak premises, then test them to see how thoroughly they account for “considered convictions” of justice. These are then adjusted against one another—changing the premises to account for previous convictions, or modifying our existing judgments (20–22). Although Rawls is not entirely clear, he does say we are not generating the principles directly, but are “altering the conditions of the contractual circumstances” (20). We are deciding on the circumstances of the original position; the principles are then generated by those circumstances. This process serves, ultimately, to remove the individual, and replace her or him with as close to a pure procedure as Rawls can construct. This is the “procedure familiar in social theory”. Decisions made will be “strictly deductive” (119). Rawls asserts that the output of these particular principles is the only possible result of his original position, as configured. Whether this is an acceptable way to proceed is a valid question; but doubting the intention distances us from the argument of the text. Rawls states that the machinery he introduces cannot be supposed in our ordinary lives. Such attempts to “simulate the original position in everyday life” will be impeded by “our various propensities and aversions” (147). *That we will fail to inhabit the original position simply by casting our minds there illustrates the need for the model.* We can move *towards* the mindset of the original position, without arriving successfully. If we view the original position incorrectly, as an environment inhabited by an actual full consciousness, then we should find our confidence in the system’s results undermined. Yet this is not the case: “But none of this affects the contention that in the origi-

nal position rational persons so characterized would make a certain decision. This proposition belongs to the theory of justice" (147).

V. MECHANISTIC MODELS IN RAWLS'S EARLY WORK

We were dismissive of Rawls connection with economic prediction above, in relation to Friedman. But there is a sense in which that needs to be revisited. Rawls, in progressing towards the formulation in *TJ*, did in some sense utilize the idea of prediction as test. In an early graduate school paper, "A Brief Inquiry into the Nature and Function of Ethical Theory", Rawls (1946) postulated that moral philosophers should search for a model which could be described as a 'reasoning machine'. These machines would be:

Systems of definitions and axioms such that when fed determinate input regarding the sorts of familiar moral choices with respect to which we can noncontroversially distinguish competent from incompetent judgments, they yield theorems or moral principles, that provide sufficient reasons for, and thereby render intelligible to us, all and only the competent judgments. (Reidy 2014, 13; much of what follows is indebted to this paper)

How predictive this all is in a Friedman sense is not completely clear, but it does have a version of 'data'—the difference between competent and incompetent noncontroversial moral judgments—which would allow a reasoning machine candidate to fail. Reidy goes on to make a connection with Frege's use of a similar 'machine' to distinguish valid inferences, and outlines a connection with the Vienna Circle and positivism. What logical positivism asserts that is helpful here is its narrowing of what counts as meaningful: either analytical statements (for example, tautologies) or synthetic statements, which can be verified or falsified by evidence (Caldwell 1982, 13). This eliminates meaning in metaphysical statements, and avoiding metaphysical options was a relative constant across Rawls's various iterations of theory.

It is interesting, and significant, to see a mechanistic design at the heart of Rawls's deliberations so early. Additional evidence from the archives is presented by Galisanka (2019). From a 1962 lecture, "Nature of Political and Social Thought and Methodology", we find: "Ideally, we want an account which reduces to zero the need for the other person to rely on his intuitive hunches as to how we will judge" (quoted in Galisanka 2019, 183; underlining in the original). In that lecture, the idea

of a machine is still present, even as the material for *TJ* is taking shape. That machine performs similarly to the much earlier 1946 version, in that the theorist should search for principles that, when built “into a machine”, would produce our own moral judgments when “given the facts and information which we regard as relevant” (quoted in Galisanka 2019, 182). One immediately noticeable change in transitioning to *TJ*, however, regards the role of information. These earlier mechanistic constructs relied on information—in 1946, more information improved results, without qualification. By 1962, the information is that “which we regard as relevant”. By the publication of *TJ*, the hugely distorting, and hence dangerous, role of information takes center stage. The original position is the response to this realization. The mechanistic function which outputs the two principles is adapted from earlier versions. The example of economics becomes crucial to the idea of rationality, and establishes parallels for how the automatic response of Rawls's procedure might work. Rawls envisions all these parts as critically interdependent:

The notions of the basic structure, of the veil of ignorance, of a lexical order, of the least favored position, as well as of pure procedural justice are all examples of [simple concepts that can be assembled to give a reasonable conception of justice]. By themselves none of these could be expected to work, but properly put together they may serve well enough. (*TJ*, 89)

Rawls's use of the phrase ‘reasoning machine’ for these earlier versions of his deliberative process makes unavoidable what should have been clear from the later work. The deliberative process should be understood as largely automatic, and mechanical. It is a model, in the sense in which we have been examining it in economics. This understanding of the deliberative position—so boldly presented in the early work—is refined and expanded in *TJ*. A somewhat lengthy quote covers several topics:

One should note also that the acceptance of these principles is not conjectured as a psychological law or probability. Ideally anyway, I should like to show that their acknowledgment is the only choice consistent with the full description of the original position. The argument aims eventually to be strictly deductive. To be sure, the persons in the original position have a certain psychology, since various assumptions are made about their beliefs and interest. These as-

sumptions appear along with other premises in the description of this initial situation. But clearly arguments from such premises can be fully deductive, as theories in politics and economics attest. We should strive for a kind of moral geometry with all the rigor which this name connotes. Unhappily the reasoning I shall give will fall far short of this, since it is highly intuitive throughout. Yet it is essential to have in mind the ideal one would like to achieve. (121)

Rawls's inability to fully implement his 'moral geometry' does not imply that we should lose track of it, as an ideal. He reminds us of this, it would seem, because he suspects this is what we might tend to do—lose track of it. There is a 'certain psychology' present as well. We might well be tempted to focus on that, and synthetically inflate its dimensions, rather than see it merely as an approximation of his actual model: moral geometry. But we should also remember how psychological assumptions were mentioned when Rawls's discussed demand theory (above). In that context they implied no attempt at verisimilitude; nor should they have such an implication here. Recall as well Mill's use of geometry in his example. There, no line as defined by mathematicians had any connection to something that might exist in the world. That does not mean geometry cannot have real-world implications. In an obvious example, geometry can aid in such things as bridge building (Hands 2001, 23). For Rawls such geometry is clearly a continuation of the automatic nature characterizing his earlier 'reasoning machine'. In these terms, what is in the original position is something, more than someone. There could be any number of these somethings, any number of original positions. ("We may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution" [*TJ*, 121].) Rawls has *constructed* the one which will output his desired principles.

It is unclear when Rawls began reading economics seriously; Reidy is working in much of the above with a paper Rawls wrote in 1946. If this predates his familiarity with economics, Rawls is establishing a number of viewpoints and working practices that he will discover dramatically overlap such ideas in economics.¹² Rawls's reasoning machine is clearly designated as such to avoid criticism for failing to resemble actual per-

¹² We do know that Rawls sat in on economics seminars his last year in graduate school, as he had finished his dissertation early. One of these was taught by Jacob Viner, who had spent the majority of his career at Chicago with Knight.

sons. It is constructed to generate output of a certain sort.¹³ It is interesting, on the one hand, to chart the evolution of Rawls's system as it changes shape and pivots on different concerns. But it is also of interest to see consistent preoccupations and solution-types visible at widely varying times. One such idea is how his system differs from intuitionism. Rawls struggles to get around intuition as a general position and fallback as early as his doctoral dissertation. There, his search for “principles [that] are like functions in mathematics” would result in principles “so precise that they are ‘mechanically followed’, or used by consciously applying the rules without appeal to intuition’ (Rawls 1950; quoted in Galisanka 2019, 79). By 1971 Rawls is still working to circumvent a position he admits “may be true”. He argues in *TJ* that his own system succeeds against intuitionism at least on one count: it narrows the scope of inquiry. In his system “we have asked a much more limited question and have substituted for an ethical judgment a judgment of rational prudence” (*TJ*, 44). Rational prudence will, suitably situated, reduce (or eliminate!) the variety of ethical positions inherent in moral judgment based on intuition. But in both accounts precision and consensus are target goals.

His argument from a linguistic parallel has a similarly precise and mechanistic feel:

There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life; it may eventually require fairly sophisticated mathematics as well.¹⁴ This is to be expected, since on the contract view the theory of justice is part of the theory of rational choice. (47)

The straightforwardly technical character of the foundation of Rawls's thought is here difficult to overlook. And rationality, at this stage at least, controls the decision-output so completely that Rawls can state the principles his pure procedure (one defined as having no set out-

¹³ “We want to define the original position so that we get the desired solution” (*TJ*, 141).

¹⁴ I cannot find anyone picking up on Rawls's references to mathematics in the literature. But the literature is large.

come) will determine. And rationality, as Rawls indicates, bonds the system's functionality all the more firmly to economics.¹⁵

We have seen that Rawls's use of models, though (perhaps) extreme and perplexing by philosophical standards, was anything but in the context of positivism and economics. Critics, however, sometimes have difficulty refraining from rushing in. Nagel, for example, criticizes Rawls's original position in terms of its excessive disconnectedness.

The model contains a strong individualistic bias, which is further strengthened by the motivational assumptions of mutual disinterest and absence of envy. These assumptions have the effect of discounting the claims of conceptions of the good that depend heavily on the relation between one's own position and that of others. (Nagel 1975, 9)

Rawls would approve of this assessment (focusing on disinterest), seeing it as a critical component of his system rather than a criticism. What Nagel here considers a too individualistic foundation to the original position is merely a blind alley. He commits the error of imagining different individuals, with differing needs and inclinations, in the original position. Not seeing Rawls's model—a term Nagel does use!—in its proper form distracts from more pertinent analysis. But even critics seeking to justify Rawls's original position often fail to see Rawls's use of models clearly. Samuel Freeman could stand for these critics. Freeman defends Rawls with the example of a mathematician. For Freeman, the mathematician need not “keep in mind particular facts about their personal lives in order to successfully” solve a problem (2007, 160). Only portions of their mental abilities—divorced from much of their personal history—need be operational. But the example puts pressure on the vulnerable point: are the ‘givens’ operating behind the veil similar to those in play for mathematics? The parallelism does not, in fact, seem to hold. Critics might respond to this defence by saying: ‘Yes, it is exactly how these situations are *not* parallel that is informative. Moral underpinnings do not have the freestanding existence that mathematical theory does. Removing the personal history building blocks that fashion our beliefs creates problems’. But this point-counterpoint argument also misses the

¹⁵ “We suppose that the parties are rational, where rationality (as distinguished from reasonableness) is understood in the way familiar from economics” (Rawls 2001, 87). Also, “moreover, the concept of rationality must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends” (*TJ*, 14).

essential character of the original position, which makes the exchange moot.

One can bypass these questions altogether of course by counter-theorizing that the entire construct of the original position is unnecessary, as Scanlon (1982) does. And it is not that all commentators have failed to follow Rawls when he discusses figures in the original position as idealizations. Nussbaum, for instance, does use the word “model”, and calls the individuals behind the veil “imaginary” (2015, 5, 17). Similarly, O’Neill recognizes the “no actual persons take part in or count in the Original Position” (2015, 60). But these recognitions still seem to stop short of the radical usage Rawls sometimes has in mind. His critics almost never discuss (or is it actually never?) his models in the mechanistic terms in which Rawls first conceived of them. This gives us perspective on arguments about how literally we are to interpret the text; that is, about its rhetorical nature. Laden (2014) speculates that Rawls utilized the language (in *TJ*) that would most clearly convince an audience of “utilitarian mathematicians and economists as well as moral philosophers”. This called for the deployment of a “formal apparatus” that was not, in fact, “essential to the argument” (67). This is an interesting viewpoint, but neglects to account for how those very formal features, designed to convince, were downplayed, or overlooked altogether. This, combined with the consistency of Rawls’s usage of mechanistic elements in the previous decades, leads us to believe that Rawls utilized this formal treatment *even though* it would disconcert. Whether we are comfortable with it or not, it forces us to be outside the system, largely to observe it rather than participate in it. Adam Smith famously exhorts us to exercise a Stoic vision of others, to value those distant from us the same as we value those near at hand. But he knows this to be impossible. We naturally favor friends and family over faceless and distant strangers. Rawls would have us view his conclusions as dispassionately and abstractly as possible—to evaluate them without consciousness of our own position, sex, race, or advantages. This is also impossible. Rawls often prefaces sentences with the word ‘ideally’. He states that when making moral decisions we often strive for the sort of objectivity and generality that the original position represents in the extreme. In that sense, we are asked to extrapolate beyond our possible experience. This distances us from the procedure, but the outcomes become proximate in the practical sphere. We can then ask: Do these outputs, emerging from this abstract cauldron, align with our deeply held beliefs? And

are they apt to be politically *useful*? These are elusive and difficult questions. But we miss them, and much else, if we get ensnared, stumbling over issues of the verisimilitude of ‘individuals’ in the original position.

A somewhat unusual parallel might prove informative. Morgan in her models book devotes a section to models as caricatures.¹⁶ Since her book is about modeling in economics, the brief look at caricatures is meant to lend some insight into what is going on with economists when they use models. Her visual example is the caricature of Louis Philippe in the middle of the nineteenth century by Charles Philipon. The caricature is that of a pear, with only a few lines to indicate that a face should be imagined on the fruit. The caricature had bite because the word for pear in French has the connotations of fathead, or dupe. And since it was the King he was caricaturing, Philipon was arraigned in court. For his defense he drew four images, the first being a portrait of the King, and the last being his pear caricature, with the intervening two being stages of abstraction. His defense was two-fold: (1) the first portrait of Louis Philippe, though quite accurate, had no indication that he was the King, and (2) really, the caricature looked like a pear, not so much like any particular person. The short story is he lost his case. But his pear won the war, as citizens across France scribbled it on buildings and fences with the obvious intention of ridiculing the monarch (Morgan 2012, 157–164).

For our purposes, we realize that the pear—while not losing touch with its origin in the face of Louis Philippe altogether—is also something *other* than a portrait of a person. It models, you could say, a single dimension of its subject—his fatheadedness. This otherness is the bridge that Rawls’s critics have trouble crossing. When one describes a person in a certain situation as ‘imaginary’, the unreality of the situation does *not* imply that we’ve given up on conceiving of the person as someone dimensionally similar to ourselves. Imagined scenarios *can* partake of sufficient realism to make them compelling. It is a leap of a different sort when a person becomes a pear, or an economic actor becomes a slot machine. Such transformations are driven by their end result; it is the insult of the pear, or the automatic nature of the slot machine, that

¹⁶ Gibbard and Varian (1978) discuss economic models as caricatures in a slightly different way. For them, when a new “feature of the world” assumes centrality in a model, “the representation of the feature is not so much an approximate description of the feature and its place in the world as it is a caricature. By that we mean not only that the approximation is rough and simple, but that the degree of approximation is not an important consideration in the design of the model” (Gibbard and Varian 1978, 673).

determines how the ‘person’ winds up in the model, not the strength of the connection back to real individuals. If economic actors are modeled as wealth maximizers (Mill’s concept), then it isn’t just that we are focusing on individuals in a market context. In that context someone might well be motivated by charity, or compassion, or any number of conflicting impulses. But the model screens those out. It aims at a *result*. Rawls maintains that his model of the original position works in a similar fashion. We begin with the outcome, and then puzzle out what dimensions of human reasoning would generate that outcome, and then isolate them. Thus, individuals behind the veil are more than (or actually less than) imagined; they are simply aspects. Can aspects alone come to conclusions about principles of social morality? Well, they had better be able to, because that is the entire motivation behind their distillation. The charge against Rawls in this area might be that those particular aspects fail in their mission, and generate something other than Rawls’s two principles. The charge cannot be that those aspects display a certain disqualifying lack of verisimilitude.

This level of abstraction in economics frequently involves assumptions about the environment in which actors operate as well. The perfection in ‘perfect competition’ obviates individual initiative. Knight believed that, in the real world, actors were searching, and in a state of becoming, without fixed preferences to strictly order their behavior. Yet his *homo economicus* model occupied the opposite extreme.

With uncertainty absent [...] it is doubtful whether intelligence itself would exist in such a situation; in a world so built that perfect knowledge was theoretically possible, it seems likely that all organic readjustments would become mechanical, all organisms automata. (Knight 1921, 268)¹⁷

In the actual world, a variety of motives and concerns can motivate individuals in identical circumstances. In such a situation, one’s economic prediction might attain Mill’s ‘most of the time’ standard. But Rawls seeks unanimity, as he tells us.¹⁸ Unanimity is a very mechanistic result.

¹⁷ We should recall here that Rawls was an exceptionally attentive reader of Knight. See footnote 5 (above).

¹⁸ “Moreover, if in choosing principles we required unanimity even when there is full information, only a few rather obvious cases could be decided. A conception of justice based on unanimity in these circumstances would indeed be weak and trivial, but once knowledge is excluded, the requirement of unanimity is not out of place and the fact that it can be satisfied is of great importance. It enables us to say of the preferred conception of justice that it represents a genuine reconciliation of interests” (*TJ*, 141–142).

There is no room for a variety of personal motives. And there is no place for a probability calculation that varies according to the participant's attitude toward risk. Rawls rules those considerations out. He preserves only those aspects of evaluation that will generate his desired outcome. In actual coalition-building, attaining unanimity has one huge virtue, and one huge cost—both obvious. The cost is convincing everyone down to the last misanthrope that your program is preferred. The advantage is there is no coercion in enforcing the policies approved—they are what everyone wants. These two costs—decision and coercion—were graphed against each other in a book by Buchanan and Tullock, that Rawls read, footnoted in *TJ*, and with the authors of which he initiated a correspondence.¹⁹ Buchanan (who advanced the constitutional perspective found in the book further on his own) and Rawls each felt their programs had significant commonality in their types of model-solutions, and in their assessment of the dangers these models helped avoid.

But how to achieve the virtues of unanimity, absent the costs? One technique which most likely won't work is simply to have a freewheeling discussion. Rawls and Buchanan share a strong joint influence in Knight, and Knight strongly opposes the notion of arriving at anything resembling what he terms 'truth' through the act of persuasion. Rawls repeats these arguments from Knight in *TJ*. Knight himself circles this problem of unanimity without making headway. The impasse for Knight is based on a lack of faith in everyday citizen thought—and this problem draws him towards a world of experts. Rawls plunges deeply in the other direction; he wants a system to output principles which determine *noncontroversial* judgments of right.²⁰ In what Rawls calls the 'science of ethics' (a phrase Knight also uses) in his early work, there is room for honest disagreement; such disagreements are in good faith, and qualify as reasonable. His system is designed to root out positions *not* held in good faith, positions whose motivation is pointedly self-interested, or based on power or class or general group-bias. Rawls needs principles he can utilize against such opinions, but the more abstract principled views he

¹⁹ The book is *The Calculus of Consent* (1962). The book was the spark to form a group—the Committee on Non-Market Decision-Making, later the Public Choice Society. Rawls was invited to, and attended, their second annual meeting. For a selection of the Rawls/Buchanan correspondence, see Peart and Levy (2008).

²⁰ I am using the word 'noncontroversial', from Rawls's early work, as a shorthand term. In *TJ* he describes the same idea more elaborately: "We can note whether applying these principles would lead us to make the same judgments about the basic structure of society which we now make intuitively and in which we have the greatest confidence" (*TJ*, 19).

seeks and these everyday muddying concerns would seem to be intertwined. Modeling is Rawls's method of disentanglement.

VI. CONCLUSION

It is possible to see models in two different ways: (1) as a world unto themselves, and (2) as a stylized version of the actual world. Each of these must have some connection to the actual world to be of interest, but their connection is not the same. Morgan sees these as distinct (Morgan 2012, 30–37). But it should be possible to see them as degrees of abstraction as well. In a sense, model builders use the level of abstraction required to order the model's world, and to secure the outcome the model was designed to examine or produce. In economics, perfect competition would be a failed model if only 'a large number' of buyers were price takers. If some really possessed bargaining power, the point of the model disappears. In other words, if the model were less extreme, and more realistic, it would become pointless. As it is, it has heuristic value, and is a graph in almost every principles textbook. It is important to keep this sort of reasoning in mind when examining Rawls's original position specification. Failure to see 'ourselves' in the original position may seem a telling criticism, but in fact such criticism points to a misunderstanding of Rawls's system. The original position has a test for success: it must output principles that coordinate with or predict our judgments of noncontroversial moral questions. In the mind of the model builder, this requirement demands a level of severity. If, for instance, the 'individual' or process in the original position were to have memory of its wealth level, this could be considered a gain in 'realism'. We would be more apt to recognize ourselves in this position. And shouldn't this greater ability to associate one's personal reality with the model construction add to our assurance about its relevance? Even more critically, we might want particular memories concerning how we formed, or solidified, some of our moral predispositions. Perhaps we feel, with these and other additions, we would gain confidence about the ability of the construct to output the same principles it did under the alternative specification. If this were so, however, we would have competing—different—theories of justice. By not selecting them, when he could have, Rawls implies these alternate models would prove less robust than his own. We must suppose he considered these 'fleshier', more accurate versions of a deliberative self/process, and found they failed to output the required principles. Why abstract more than one

needs to? The process he gives us is exactly the process that performs as required—meeting, as it were, its design brief. And we have every textual reason to suppose, despite it having ‘a certain psychology’, that the decision-process in the original position can most successfully be understood as a sort of mechanism. To ignore this, and desire that the process doing this work were more like us, is to misunderstand Rawls’s program in a fundamental way.

Morgan suggests that philosophers have problems with the concept of models for a number of (well-founded) reasons. She mentions particularly that there is “concern about the status of the representation” (2012, 33). What does the model ‘mean’ if it is something other than the world? Sugden discusses the lure of instrumentalism, where “the ‘assumptions’ of a theory, properly understood, are no more than a compact notation for summarizing the theory’s predictions; thus the question of whether assumptions are realistic or unrealistic does not arise” (2002, 117). For Rawls, this would seem to imply looking to applying principles to the noncontroversial moral situations. Or, as Rawls phrases it in *TJ*, the output must align with certain convictions. And “these convictions are provisional fixed points which we presume any conception of justice must fit” (*TJ*, 20). This total scrapping of the model’s link to reality—except through the predictive tie—makes the terms of description in models somewhat misleading. In economics as well, calling the units in the model ‘individuals’ necessarily creates a tension. We have a tendency—across disciplines—to imagine them, at least in certain ways, like ourselves. We want to have some grip on the nature of their ‘status of representation’. Because we intend the predictive outcome of the model to apply to individuals, the natural inclination is to envision the model in individual terms as well. Rawls, like economists, necessarily blurs our picture by talking of the model both as a complete abstraction, and as some version of an individual. We focus on this latter representation not just because it is familiar, but also because it gives us another way to grasp the theory. Models as total abstractions are the creatures of their creator; they have purpose without dimension. And in the sense that they are a black box, they resist our criticism. As critics, we find this an obstruction. We want to choose our interpretation of what the model is, in order to bypass this impediment.

Perhaps we need to be more cautious than this, however. It is simply being argued that, in light of a natural tendency to dress our models as people, we should be cognizant of how this can work against, rather

than aid, our better understanding. Rawls points on several occasions to the tension we're discussing; this is from *Political Liberalism*:

As a device of representation, its *abstractness* invites misunderstanding. In particular, the description of the parties may seem to presuppose a particular metaphysical conception of the person [...].

I believe this to be an illusion caused by not seeing the original position as a device of representation. The veil of ignorance, to mention one prominent feature of that position, has no specific metaphysical implications concerning the nature of the self; it does not imply that the self is ontologically prior to the facts about persons that the parties are excluded from knowing. (*PL*, 27)

Rawls then introduces the comparison with role playing. When acting the role of Macbeth or Lady Macbeth, we shouldn't be thinking we *are* plotting nefariously in Scotland! We must keep in mind what the exercise is about: "Trying to show how the idea of society as a fair system of social cooperation can be unfolded so as to find principles specifying the basic rights and liberties and the forms of equality most appropriate to this cooperating, once they are regarded as citizens, as free and equal persons" (27).

In conclusion, it might be fitting to examine another thinker who has strongly influenced both philosophy and economics—Thomas Hobbes. If we can talk about Hobbes's ([1651] 1968) 'state of nature' as a model, it offers some contrast with Rawls's. In Hobbes, it is the situation that is modeled, but you are supposed to imagine yourself in that situation, fully formed. Hobbes's model also fails the realism test (as, we have been arguing, almost every model does). In the real world, individuals bunch into groups or tribes, and the warfare is external. Internally, within-group, there is relative harmony. Hence Hobbes's model is 'incorrect'. Rawls's model would seem to ask less of us—we don't need to imagine a new exterior environment. But in certain ways it asks much more. It asks, at the least, for us to enter a mode in which we are not fully ourselves. More accurately, though, it asks of us to become a slice of reasoning behavior, and not really be ourselves at all. Models, to function, can be extreme. Rawls requires his system to generate very specific and powerful conclusions. And to accomplish this his model, despite its nominal parallels with merely sequestered individuals, is at core more extreme than most.

REFERENCES

- Audard, Catherine. 2007. *John Rawls*. Montreal: McGill-Queen's University Press.
- Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, MI: The University of Michigan Press.
- Caldwell, Bruce. 1982. *Beyond Positivism: Economic Methodology in the Twentieth Century*. London: George Allen & Unwin.
- Coker, David C. 2021. "Rawls and Knight: Connections and Influence in *A Theory of Justice*." *Research in the History of Economic Thought and Methodology* 39C: 77–98.
- Freeman, Samuel. 2007. *Rawls*. London and New York, NY: Routledge.
- Friedman, Milton. (1953) 1966. "The Methodology of Positive Economics." In *Essays in Positive Economics*, 3–43. Reprint, Chicago, IL: The University of Chicago Press (Phoenix Books).
- Galisanka, Andrius. 2019. *John Rawls: The Path to A Theory of Justice*. Cambridge: Harvard University Press.
- Gibbard, Allan, and Hal Varian. 1978. "Economic Models." *Journal of Philosophy* 75: 664–677.
- Hands, D. Wade. 2001. *Reflection Without Rules: Economic Methodology and Contemporary Science Theory*. Cambridge: Cambridge University Press.
- Hobbes, Thomas. (1651) 1968. *Leviathan*. Reprint, Harmondsworth: Penguin Books.
- Knight, Frank H. (1921) 1971. *Risk, Uncertainty, and Profit*. Reprint, Chicago, IL: The University of Chicago Press.
- Knight, Frank H. 1935. *The Ethics of Competition and Other Essays*. London: George Allen & Unwin, Ltd.
- Laden, Anthony Simon. 2014. "Constructivism as Rhetoric." In *A Companion to Rawls*, edited by Jon Mandle, and David A. Reidy, 59–72. Malden, MA: Wiley Blackwell.
- Mill, John Stuart. (1967) 2006. *Collected Works of John Stuart Mill, Vol. IV: Essays on Economics and Society 1824–1845*. Reprint, Indianapolis, IN: Liberty Fund Press.
- Morgan, Mary S. 2012. *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1975. "Rawls on Justice." In *Reading Rawls: Critical Studies of A Theory of Justice*, edited by Normal Daniels, 1–15. New York, NY: Basic Books, Inc.
- Nussbaum, Martha C. 2015. "Introduction." In *Rawls's Political Liberalism*, edited by Thom Brooks, and Martha C. Nussbaum, 1–56. New York, NY: Columbia University Press.
- O'Neill, Onora. 2015. "Changing Constructions." In *Rawls's Political Liberalism*, edited by Thom Brooks, and Martha C. Nussbaum, 57–72. New York, NY: Columbia University Press.
- Peart, Sandra J., and David M. Levy, eds. 2008. *The Street Porter and the Philosopher: Conversations on Analytical Egalitarianism*. Ann Arbor, MI: The University of Michigan Press.
- Rawls, John. 1946. "A Brief Inquiry into the Nature and Function of Ethical Theory." Box 3, Accession 14990, Rawls Archive, Pusey Library, Harvard University, Cambridge.
- Rawls, John. 1950. "A Study in the Grounds of Ethical Knowledge: Considered with reference to Judgements on the Moral Worth of Character." PhD dissertation, Princeton University.

- Rawls, John. 1962. "Nature of Political and Social Thought and Methodology." Lecture IX, 2i, HUM 48, Box 35, Folder 10, The Papers of John Rawls, Harvard University Archives, Pusey Library, Harvard University, Cambridge.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: The Belknap Press of Harvard University Press.
- Rawls, John. (1980) 1999. "Kantian Constructivism in Moral Theory." In *Collected Papers*, edited by Samuel Freeman, 303–358. Reprint, Cambridge: Harvard University Press.
- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge: The Belknap Press of Harvard University Press.
- Rawls, John. 2005. *Political Liberalism: Expanded Edition*. New York, NY: Columbia University Press.
- Rawls, John. 2007. "Of the Original Contract." In *Lectures on the History of Political Philosophy*, edited by Samuel Freeman, 103–121. Cambridge: The Belknap Press of Harvard University Press.
- Reidy, David A. 2014. "From Philosophical Theology to Democratic Theory." In *A Companion to Rawls*, edited by Jon Mandle, and David A. Reidy, 9–30. Malden, MA: Wiley Blackwell.
- Sandel, Michael J. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Scanlon, Thomas M. 1982. "Contractualism and Utilitarianism." In *Utilitarianism and Beyond*, edited by Amartya, and Bernard Williams, 103–128. Cambridge: Cambridge University Press.
- Spiegel, Henry William. 1983. *The Growth of Economic Thought*. Durham: Duke University Press.
- Sugden, Robert. 2002. "Credible Worlds: The Status of Theoretical Models." In *Fact and Fiction in Economics: Models, Realism, and Social Construction*, edited by Uskali Mäki, 107–136. Cambridge: Cambridge University Press.
- Taylor, Robert S. 2014. "The Priority of Liberty." In *A Companion to Rawls*, edited by Jon Mandle, and David A. Reidy, 147–163. Malden, MA: Wiley Blackwell.
- Varian, Hal R. 1984. *Microeconomic Analysis*. 2nd edition. New York, NY: W. W. Norton & Company.

David C. Coker holds a BA in English Literature from Amherst College and a PhD in Economics from George Mason University, and is currently teaching History of Thought at the University of Maryland, Baltimore County. He has published work on James Buchanan, Frank Knight, and John Rawls.

Contact e-mail: <dccoker@mindspring.com>