ECONOMICS TO SOCIAL PHILOSOPHY: THE FOUNDATIONS OF JOHN RAWLS'S CONTRIBUTION

by

David C. Coker

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Economics

Committee:

_____	Director
_____	
_____	
_____	Department Chairperson
_____	Program Director
_____	Dean, College of Humanities and Social Sciences
Date: _____	Spring Semester 2021 George Mason University Fairfax, VA

Economics to Social Philosophy: The Foundations of John Rawls's Contribution

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

David C. Coker
Master of Arts
George Mason University, 2018
Bachelor of Arts
Amherst College, 1982

Director: David M. Levy, Professor
Department of Economics

Spring Semester 2021
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

Primary thanks go to David Levy, who is largely responsible for my getting back into economics. It was reading his *How the Dismal Science Got its Name* that stimulated me to make that initial phone call. In addition to his example and encouragement, he also lit the Rawls fire by providing scans of Rawls's annotations of Frank Knight's essays in *The Ethics of Competition and other essays.* Much of what is here presented follows in one way or another from that experience. Additional thanks go to Erik Angner, who took an interest at a critical juncture, and pushed me towards conferences early in my graduate school experience. Ross Emmett was also a positive force in the early stages of this project. My wife, Cathy, provided the space for me to do this. She also read everything I wrote, and kicked me when needed.

TABLE OF CONTENTS

Page

Abstract v

1. Rawls and Knight: Connections and Influence in *A Theory of Justice* 1

I. Introduction 1

II. The Market as Inadequate Distributional System 6

III. Political Options 13

IV. Discussion, Consensus, and the Original Position 19

V. Conclusion 28

2. Modeling in Rawls: The Original Position 33

I. Introduction 33

II. 35

III. 40

IV. 44

V. 57

3. Informational Restrictions in Rawls's Original Position: Economics and the Inconsistent Plans Literature 64

I. Introduction 64

II. Background 75

III. Informational Restrictions 78

IV. Rationality 87

V. Conclusion 96

References 100

ABSTRACT

ECONOMICS TO SOCIAL PHILOSOPHY: THE FOUNDATIONS OF JOHN RAWLS'S CONTRIBUTION

David C. Coker, Ph.D.

George Mason University, 2021

Dissertation Director: Dr. David M. Levy

These essays are an attempt to understand Rawls's use of economics in developing his system in *A Theory of Justice.* The intensive use of and reference to economic theory, I will argue, is deeply involved with how the theory was originally conceived. The critical response to Rawls has thus far mostly ignored this line of thinking. For instance, Pogge's reference to Rawls building an early version of his original position based on an idea in Knight (as relayed to Pogge in personal conversation) has resulted in almost no connection between Rawls and Knight in the literature. Rawls's extensive use of parallels with economics is key to how the system functions. Strands of analysis connected to the idea of a thought experiment, to decisions made behind the veil of ignorance, and to the system relying on moral impulse as motivation are all tangential to what actually drives the system. Rawls in fact utilizes, at base, a rather severe form of modeling. This sets up an enormous contrast with utilitarianism, where increases in levels of information and of

benevolence are both central ingredients. Neither benevolence nor full information are at the center of the process for Rawls. There is an automaticity at the hinge point in Rawls that has baffled a large portion of the philosophical community. These papers hope to explain some of these elements.

1. RAWLS AND KNIGHT: CONNECTIONS AND INFLUENCE IN *A THEORY OF JUSTICE*

For [Rawls], as for the Marxists, positivists, and Utilitarians, moral systems are creations of human societies, designed to solve problems that arise when people live together.

--Samuel Fleischacker, *A Short History of Distributive Justice*[1]

No doubt we all agree that extremes of wealth and poverty are unjust -- especially when they do not correspond with personal effort or sacrifice -- and are bad in other ways. The question is, what can we do about it? Can the rules of the economic game be so changed that the winnings, symbolic and real (and the former are not much inferior in importance), will accord better with some accepted or defensible criterion of justice? And can it be done without wrecking the game itself, as a game, and as a producer of the fruits on which we all live?

-- Frank Knight ("The Role of Principles in Economics and Politics.")[2]

I. Introduction

While Rawls was working on the papers that would lead to *A Theory of Justice*, the sorts of questions he was attempting to address were not among the leading issues in contemporary philosophy. Partly to fill this void, Rawls turned to economics. There he found, apparently, much that was germane to his inquiry. One short passage will serve as

[1] Fleischacker 2004, 110.
[2] Knight 1951, 20.

an indication. Part of his on-going criticism of utilitarianism, this passage occurs in the early essay, "Justice as Fairness", where he is arguing for the distinctiveness of the title concept,

> For one thing, that the principles of justice should be accepted is interpreted as the contingent result of a higher order administrative decision. The form of this decision is regarded as being similar to that of an entrepreneur deciding how much to produce of this or that commodity in view of its marginal revenue ….[3]

This reference to marginal revenue seems jarring in a philosophical paper. Yet this low key economic reference is indicative of a more serious interaction with economics, that was to culminate in many ways in *Justice*. There are over twenty separate economists footnoted in *Justice*, but it isn't quantity alone that is of interest. Much of the argument is approached economically, and markets serve as a benchmark for certain philosophical concepts, up to and including dimensions of fairness.

This paper will focus on economic issues primarily through the lens provided by a particular economist – Frank Knight. Knight's book, *The Ethics of Competition and other essays*[4], was read by Rawls independently, probably late in his graduate studies. It was part of a general attempt to inform himself on economics. Having finished his thesis a year early at Princeton, Rawls spent much of that extra time reading economics and attending economic seminars (one was taught by Jacob Viner, who spent much of his academic life

[3] Rawls 1958, 186 (and in *Collected Papers,* 65).
[4] *Ethics* is now back in print, but my page references will be to the journals where the articles originally appeared, *and* to the original 1935 edition which Rawls annotated. The exception is "Economic Theory and Nationalism", which first appeared in the collection, so will be noted simply as Knight 1935.

at the University of Chicago with Knight). But Knight's book may have been particularly important. Rawls credits one of the essays with generating his focus on principles justified "by reference to an appropriately formulated deliberative procedure".[5] This alone would be reason to integrate Knight into one's interpretation of *Justice*. But a closer look at the essays in *Ethics* shows this to be a small part of the influence. There are a host of areas where Knight's ideas seem to be present, beyond the handful that generate entire arguments and are footnoted. This suspicion can now be reinforced, through the presence of Rawls's annotated copy of *Ethics*.[6]

The bulk of this paper will attempt to show, through parallels between Knight's essays in *Ethics* and Rawls's *Justice*, that not only do Rawls's arguments track Knight's closely, but that they *probably originated there as well.* As exciting as this in its own right, it also has certain implications. If a basket of ideas in Rawls can be tied to Knight, it shifts our perspective about the periods in which they are best intellectually contextualized. Since Knight's essays were written in the 1920's and 1930's, some contextual emphasis shifts from post- to pre-WWII. Thus the Socialist Calculation Debate has a newfound relevance. The liberal program, in its nineteenth century guise, was under a different sort

---

[5] "During this period [end of graduate school and the post-doc at Oxford], Rawls began developing the idea of justifying substantive moral principles by reference to an appropriately formulated deliberative procedure. He said that the inspiration for this idea may have come from an essay by Frank Knight, which mentions the organization of a reasonable communicative situation ("Economic Theory and Nationalism" [appeared originally in the collection *The Ethics of Competition and other essays*]). Rawls's initial idea was that the participants should deliberate independently of one another and forward their proposals for moral principles to an umpire. As with later versions of the original position, Rawls was hoping that he could derive substantive results from an exact and elaborately justified specification of a hypothetical situation – that is, without having to implement a procedure with actual participants." (Pogge 2007, 16-17)

[6] Scans of Rawls' annotations in Knight's *The Ethics of Competition and other essays* courtesy of David Levy, George Mason University.

of attack than it was post-War. Because more economic options were intellectually "on the table", defending the market was almost as radical as attacking it. Democracy, and voter competence, were also frequently seen as less reliable than reliance on experts.[7] Knight and Rawls part company on the expert question. But in Knight's intellectual milieu, democratic process as a solution to moral social problems was at one of its lowest ebbs. Each therefore faces the need to adjust the deliberative situation, to refine it in some way. They do this differently, but each is clear on what the individual needs to be isolated *from*. Here, Rawls is observing Knight's reasoning closely.

This also has implications for what might be called Rawls's ideological location. Rawls's work incorporates an enormous range and variety of influences. Critics have frequently attempted to parse this variety by seeing it along an ideological line. The assumption – that Rawls has a particular ideological position – is frequently followed by attempts to pin that position down, and use it as a reference to bring other elements into conformity. Rawls's use of economics in his system is usually assumed to align with the "right-leaning" elements. Incorporating Knight's influence disrupts this approach. Knight has a number of vehement beliefs which range considerably across the ideological spectrum. And Knight's critics, like Rawls's, found it difficult to categorize those beliefs. And many of the confounding tensions in Rawls's thought should be linked to ideas in Knight. In his talk at the 1973 AEA, Rawls asserts that he rejects preferences as static, a view (preferences as static) that is usually ascribed to economists.[8] But Knight argues in a

[7] For an excellent survey of these decades and their intellectual trends in social science see Purcell 1973.
[8] Quoted in Forrester, 125-6.

number of places, at length, that preferences or wants are in a continual process of formation, and that it is the shaping of these wants, rather than the wants themselves, that is most crucial. What we most want is, in his phrase, "better wants". Rawls marked these passages. On a more contentious topic – redistribution – we see Knight in the quote above as giving support to the unfairness of a laissez-faire system ("No doubt we all agree that extremes of wealth and poverty are unjust"). Knight is unsure even of capitalism's survival. On what might be considered the other side of the ideological spectrum, however, Knight and Rawls recognize what the market offers, and how attempts to interfere in its workings are counterproductive.[9] Many of Knight's ideas, here and analyzed below, were discovered in Rawls's graduate school period, and emerge intact more than twenty years later in *Justice.* They represent an essential basket of often conflicting ideas (or at least ideas in tension) that Rawls found congruent with his earliest impulses in social theory. But they also represent a body of ideas that were *already conjoined*. Knight's influence in these terms helps reconceive Rawls less ideologically, less in an either/or light.

This paper will take three arguments centered on justifying the original position, and examine them in relation to ideas set out by Knight. The first section will be centered on markets, in both their theoretical and practical dimensions. The second section will focus on the political system as a possible alternative to Rawls' constructs. And a third section will address the nature of deliberation behind, and in front of, the veil of

[9] Numerous examples could be cited. One from "Economic Theory and Nationalism" is "One of the main advantages of the enterprise system has been its promotion of progress, to a degree undreamed of in any other region of history" (p.310 in the *Ethics* collection). Rawls underlined "enterprise system" and "promotion of progress", and double checked in pencil in the margin.

ignorance. Rawls and Knight take on similar questions: what is the value of the market, both practically and *analytically*, and what moral value does the market offer in its workings and outcomes. What links exist, if any, between markets and moral philosophy generally? These are difficult questions, and their answers are not always positive. Knight's influence on Rawls, until now, has not been widely noted.[10] But reading *Justice* with Knight in mind brings additional emphases to the text, and makes even more substantial the interconnectedness of ideas in Rawls's complex work.

II. The Market as Inadequate Distributional System

In the beginning of *A Theory of Justice*, Rawls specifies that the concept of justice applies (for his purposes) to the "basic structure of society", and the manner in which its institutions "distribute fundamental rights and duties and determine the division of advantages from social cooperation." The four institutions he lists in the early pages are: 1) those insuring the legal protection of freedoms, 2) competitive markets, 3) private property, and 4) the monogamous family.[11] Two of the four are central to economic functioning in a market system. Rawls' approval of markets, and construction of his theory to accommodate them, is usually passed over with a brief sentence by other

[10] This omission is easily seen. In the following: *A Companion to Rawls,* Jon Mandle, ed. (Wiley-Blackwell, 2013); *Rawls's 'A Theory of Justice': An Introduction*, Jon Mandle (Cambridge University Press, 2009); *Rawls: 'A Theory of Justice' and Its Critics* by Chandran Kukathas (Stanford University Press, 1990); *The Cambridge Companion to Rawls,* Samuel Freeman, ed. (Cambridge University Press, 2002); *Reading Rawls: Critical Studies of 'A Theory of Justice'* , Norman Daniels, ed. (Basic Books, 1974); and *Why Political Liberalism?: On John Rawls's Political Turn* by Paul Weithman (Oxford University Press, 2013), there is, between them, a single reference to Knight in their indices. More recently, however, Knight is beginning to appear. See, for example, Levy and Peart (2017), and Forrester (2019).

[11] Rawls 1971*,* 7.

theorists. It is not the preservation of markets that sparks interest for commentators; it is how markets form the initial stage to what seems of greater interest: the difference principle. Samuel Freeman, in his excellent study, is representative: "it is against a background of market allocation of factors of production that Rawls assumes that the difference principle will work best to advance the position of the worst-off within a modern economy."[12] Discussions of markets would seem to offer no extra insight into the central issue of the difference principle, or into Rawls' structural set-up in general.

> It is surprising, therefore, when we encounter passages such as the following:
>
> The ideal scheme sketched in the next several sections makes considerable use of market arrangements. It is only in this way, I believe, that the problem of distribution can be handled as a case of pure procedural justice. Further, we also gain the advantages of efficiency and protect the important liberty of free choice of occupation.[13]

Rawls here uses the terms "efficiency" and "liberty" in the same sentence, and this is, in microcosm, the tension and balance of his argument throughout the book. Ideas that dominates discussion *around* Rawls tend to be focused on how the philosophical ideas in isolation relate to one another and are justified. Yet Rawls' *own* presentation generally presupposes and analyzes notions of efficiency as an abstract starting point, and explores the degree to which concepts of value are compatible with them. This tension – between efficiency and values – we will see as identical with the argument-structure used by Knight.

---

[12] Rawls 1971, 104.
[13] Rawls 1971, 274.

Rawls does a great deal more than mention markets and efficiency in passing. He applauds the market for, under certain conditions, making possible a Pareto efficient distribution of goods and choice of productive methods by firms.[14] Also,

> A further and more significant advantage of a market system is that, given the requisite background institutions, it is consistent with equal liberties and fair equality of opportunity.[15]

And somewhat later,

> Moreover, a system of markets decentralizes the exercise of economic power.[16]

These are remarkable assertions. Market structures not only diffuse concentrations of power, they even serve to embody notions of efficiency and fairness. And the power-defusing component will be at the center of critical arguments in both Knight and *Justice* (developed below). Further, as an "ideal conception", a perfectly competitive market benchmark "may then be used to appraise existing arrangements and as a framework for identifying the changes that should be undertaken."[17]

However, given that abstract markets possess all these virtues, from what does the tension arise? The tension comes in trying to make compatible market *results* and ethical justification. It is a critical part of Rawls' theory that he doesn't approve of market distribution as a just outcome – these are the "changes that should be undertaken" that he hopes his theory will "identify" in the previous quotation. Why then keep the market

---

[14] "I assume in all interpretations that the first principle of equal liberty is satisfied and that the economy is roughly a free market system, although the means of production may or may not be privately owned." (Rawls 1971, 66) Also, for the Pareto principle, Rawls footnotes Buchanan 1962, and Buchanan and Tullock 1962.
[15] Rawls 1971, 272.
[16] Ibid.
[17] Ibid.

process, and praise and utilize it as a theoretical benchmark? Rawls in a sense wishes to jettison the bathwater, but save the baby. The market has real-world productive value, and presents theoretical benchmarks for efficiency and value. To dismiss the market as a whole because of its failings in terms of moral-distributional outcomes is short-sighted, and logically flawed. Knight is explicit making the point:

> It is a common assumption – for which the exponents of the "productive theory" are partly responsible – that productive contribution is an ethical measure of desert. This has improperly tended to bring the theory itself, as a causal explanation of what happens in distribution, into disrepute; because those who are misled into accepting the standard, but cannot approve of the result realized, react by attacking the theory.[18]

According to Knight, then, it is a mistake to dismiss market theory because one doesn't approve of final market distributional outcomes. The two should be viewed as distinct. This is precisely how Rawls sees the market.[19] He rejects market outcomes as just, while accepting the market system as the preferred method for allocating and organizing economic resources, and insuring certain dimensions of freedom. He isn't "seduced" into dismissing markets altogether because they fail, despite their virtues, to generate what he considers just distributional outcomes.

Knight also believes that market outcomes lack ethical significance, and makes a

---

[18] "The Ethics of Competition", 596 in the QJE, and 54 in the *Ethics* collection.

[19] In chapter 2, Rawls similarly argues: "In view of these remarks we may reject the contention that the ordering of institutions is always defective because the distribution of natural talents and the contingencies of social circumstance are unjust, and this injustice must inevitably carry over to human arrangements." (Rawls 1971, 102)

detailed list of why this is so. From Knight's list of reasons, some of the ones which are prominent for Rawls[20] also are: 1) the "product or contribution" is measured in price, which does not correspond closely with "ethical value or human significance", 2) income goes to owners, not factors of production, and "can in no case have more ethical justification than has the fact of ownership. The ownership of material or productive capacity is based upon a complex mixture of inheritance, luck, and effort, probably in that order of relative importance", 3) "the value of any service or product varies from zero to an indefinite magnitude, according to the demand. It is hard to see that even when the demand is ethical, possession of the capacity to furnish services which are in demand, rather than other capacities, constitutes an ethical claim to a superior share of the social dividend, except to the extent that the capacity is itself the product of conscientious effort", 4) a similar argument for scarcity, and 5) a similar view of competence.[21] Thus for Knight the efficiency virtues of the market fail to carry over to questions of ethical worth of distributional outcomes.[22]

---

[20] On the page before this list (54), Rawls has written in the margin of his copy, in red ink (which he used for special emphasis): "That productive contribution has little ethical significance". (David Levy Photostats of Rawls' copy of *Ethics*)

[21] Pages 597-600 in the *QJE,* 55-57 in the *Ethics of Competition* collection. This entire section is extremely heavily marked in Rawls's copy of *Ethics.*

[22] Or, as Knight argues elsewhere, markets resembling the idealized form simply don't exist in the real world. So why tie a political/moral theory to them? Rawls asks and answers this question: "It may be objected to the preceding account of the common sense precepts and to the idea of pure procedural justice that a perfectly competitive economy can never be realized. Factors of production never in fact receive their marginal products, and under modern conditions anyway industries soon come to be dominated by a few large firms. Competition is at best imperfect and persons receive less than the value of their contribution, and in this sense they are exploited. The reply to this is first that in any case the conception of a suitably regulated competitive economy with the appropriate background institutions is an ideal scheme which shows how the two principles of justice might be realized. It serves to illustrate the content of these principles, and brings out one way in which either a private-property economy or a socialist regime can satisfy

Rawls walks this same line of argument (see previous footnote), beginning with efficiency, and noting the limitations of the market system when it comes to "just" outcomes:

> Now it is natural to try out the idea that as long as the social system is efficient there is no reason to be concerned with distribution. All efficient arrangements are in this case declared equally just. Of course, this suggestion would be outlandish for the allocation of particular goods to known individuals. No one would suppose that it is a matter of indifference from the standpoint of justice whether any one of number of men happens to have everything. But the suggestion seems equally unreasonable for the basic structure.[23]

This argument by Rawls follows an Edgeworth box-type example, where the possible Pareto-efficient set potentially includes one individual having all of both goods. But Rawls begins, as does Knight, with a baseline of a theoretically efficient system, and sees how closely it can or cannot approximate a system that deals adequately with values. And some of Rawls' particular arguments reflect those of Knight above. Rawls is sensitive to the nature of "supply" remaining consistent, while demand shifts. How can *moral* deservingness be determined purely by the activities and desires of others?

The principles of justice that regulate the basic structure and specify the duties

---

this conception of justice. Granting that existing conditions always fall short of the ideal assumptions, we have some notion of what is just. Moreover we are in a better position to assess how serious the existing imperfections are and to decide upon the best way to approximate the ideal." (Rawls 1971, 309)

[23] Rawls 1971, 71. Here we should note a comment by Lyons: "It is unfortunate, therefore, that Rawls merely claims, without supporting argument, that distributions flowing from natural or social contingencies alone are arbitrary from a moral point of view" (Lyons 1974). Lyons misses the arguments from Knight.

> and obligations of individuals do not mention moral desert, and there is no tendency for distributive shares to correspond to it.
>
> This contention is borne out by the preceding account of common sense precepts and their role in pure procedural justice (sec. 47). For example, in determining wages a competitive economy gives weight to the precept of contribution. But as we have seen, the extent of one's contribution (estimated by one's marginal productivity) depends upon supply and demand. Surely a person's moral worth does not vary according to how many offer similar skills, or happen to want what he can produce. No one supposes that when someone's abilities are less in demand or have deteriorated (as in the case of singers) his moral deservingness undergoes a similar shift. All of this is perfectly obvious and has long been agreed to.[24] (Rawls cites Knight for this section.)

Rawls elsewhere looks favorably upon worker-owned enterprises. And he attacks marginal product as a moral yardstick, arguing that since it depends on supply and demand, the moral connection is severed. "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."[25]

It would appear that, finding a parallel concern in Knight to proceed from concepts of efficiency and market, to concepts of ethical social value, Rawls follows many of the

---

[24] Rawls 1971, 311.
[25] Rawls 1971, 308. And in this section, close to Knight's ideas as it is, Rawls does not footnote Knight.

arguments and assessments Knight lays out in his earlier theorizing in *Ethics*. As parallel as their paths are, however, Rawls intends to reach a very different destination. Knight winds up on a vaguely pessimistic note: perhaps capitalist/market structures are not fated to survive. And discussion, requiring dispassionate experts to reach sound moral conclusions, is not precisely envisioned, or even anticipated. Rawls, though, takes the inequalities of market outcomes as a call for both redress, and extensive reconfiguring of the choice environment. The redress is centered on maximin and the difference principle. And reconfiguring the choice environment – avoiding self-interested choices that are further distorted by economic power inequalities – is to locate decisions away from those distortions, behind the veil of ignorance. In the next section, further support will be found for the original position concept, in the problems and issues that characterize real-world political activity.

III. Political Options

Rawls' argument for the veil has been challenged as unworkable from a variety of angles, but it has, for some critics, also been viewed as entirely dispensable. T.M. Scanlon is one of the foremost of these, not only because he outlines a competing procedure, but also because he is, like Rawls, a contractarian (though his term for such a system is "contractualism"). Scanlon argues an individual might endorse principles because they are judged to be ones "he could not reasonably reject whatever position he turns out to occupy…."[26] Scanlon's option simplifies Rawls' system, and has the additional advantage

[26] Scanlon 1982, 124. For a more complete explanation of his position, see Scanlon, 1998.

of projecting it more successfully into the real world. For Scanlon, one might assume, the principles under examination lie at a deeper level than the "interests" which Rawls is anxious to bypass, and so those interests would not be a hindrance to agreement, even outside the veil of ignorance. This appears a plausible position, and objection. And from it the temptation would be to see if real-world decision environments might serve to achieve Rawls' ends, without the complex machinery of the original position and veil of ignorance. What of, for instance, the political process? If agents can reach their deepest principles with their interests and social position still apparent to them, what is to prevent some form of political choice from replacing the intricate choice conditions Rawls' feels are required?

Rawls' arguments are both general, as relates to decisions attempted in "everyday life", and particular, as regards decisions attempted in the political process. His arguments against everyday decision-making are well-understood, so a brief rehearsal of them here will suffice. First, it is against a stringent standard – that of unanimity – that Rawls objections must be understood. Scanlon specifies "reasonable" rejection as a standard. But for Rawls the possibility of a person's interests trumping their reasonableness is eminently possible.

> Of course, when we try to simulate the original position in everyday life, that is, when we try to conduct ourselves in moral argument as its constraints require, we will presumably find that our deliberations and judgments are influenced by our special inclinations and attitudes. Surely it will prove difficult to correct for our various propensities and aversions in striving to adhere to the conditions of this

> situation. But none of this affects the contention that in the original position rational persons so characterized would make a certain decision. (Rawls 1971, 147)
>
> The veil of ignorance makes possible a unanimous choice of a particular conception of justice. Without these limitations on knowledge the bargaining problem of the original position would be hopelessly complicated. Even if theoretically a solution were to exist, we would not, at present anyway, be able to determine it. (Rawls 1971, 140)

For Scanlon, there would be no "bargaining problem". For Rawls there is. But Rawls characterizes the problem as running deeper than some form of topical selfishness. And for this he borrows from Knight a discussion on the problems of power.

We have seen in the previous sections some of Knight's objections to market activity as it is actually configured in the world. And what characterizes motivation in economic activity, and the resultant differences in levels of power available to actors, carries over completely into the political world. Politics turns out to be a dead end. Its difference from its ideal is seen as even greater than that difference for the market: "The main error on the political side, in the theory of liberalism as expounded by its advocates, is that competitive politics is not better than economics in this regard, but definitely worse." (*Ethics*, 296; checked, underlined, and double-margin marked in Rawls's copy) Returning to his theme of the ideal market being "atomistic", it is not surprising to find Knight following the same line analyzing politics. Ideal political interaction should also be atomistic — direct democracy on a small scale allows each participant a vital place. This contrasts strongly with reality, as it did in his analysis of the market. Again, the quest for

power finds some "contestants" comfortably, and increasingly, ahead of others. It is a little remarked feature of Knight's analysis that those gaining power advantages preserve and increase those advantages. In "Economic Theory and Nationalism", Knight glumly asserts, "As no one needs to be told, the realities in both business and politics have been very different from these ideals. … And the main weakness is the same in both cases, as compared with an ideal system in which 'each should count for one and none for more than one'; it lies in the natural cumulative tendency toward inequality in status, through the use of power to get more power."[27] (underlined in Rawls's text; from "it lies", much underlined twice, in pencil and then red ink, with red ink margin emphasis as well) In case we are in any doubt, Knight continues on the next page, "Thus liberal economics and liberal politics are at bottom the same kind of 'game'. The fundamental fact in both is the moral fact of rivalry, competitiveness, and the interest in power."[28] For Knight the metaphor of the game is central[29]: if players are more concerned to win than they are to preserve the game itself, then societal structures are threatened.

Rawls picks up on these assessments, and footnotes Knight in his own elaboration of these points. When discussing government specifically, Rawls describes the purpose of the "distributive branch". The focus is exactly Knight's focus: "The purpose of these levies and regulations is not to raise revenue (release resources to government) but gradually

---

[27] "Economic Theory and Nationalism" (Knight 1935), 296. One of Knight's students put it this way: "the deepest contradiction in Knight's view of human society: on the one hand, he regarded individual freedom as a basic value, and recognized that representative democracy was the only way in which a large society of free individuals could govern itself; on the other, he had basic misgivings about the actual workings of the democratic process -- and was accordingly deeply pessimistic about its future." (Patinkin, 807).

[28] Knight 1935, 297.

[29] Rawls also uses the game metaphor in this way.

and continually to correct the distribution of wealth and to prevent *concentrations of power* detrimental to the fair value of political liberty and fair equality of opportunity."[30] (emphasis added) The importance of power imbalances, highlighted in Knight's text by Rawls, find their way into Rawls' own argument intact.

> It is these institutions [guaranteeing fairness] that are put in jeopardy when inequalities of wealth exceed a certain limit; and political liberty likewise tends to lose its value, and representative government to become such in appearance only. The taxes and enactments of the distribution branch are to prevent this limit from being exceeded. (Rawls 1971, 278)

And a similarly bold statement in Knight:

> Consequently, under individualistic freedom, and under the condition that men want more wealth, for whatever reason, it will be used to get more, giving rise to a cumulative growth of inequality. Two further consequences follow in turn: (a) With "gross" inequality in the distribution of wealth among individuals, all ethical defences of freedom lose their validity; and (b) the automatic system of control (market competition) breaks down, for competition requires a large number of units, every one of negligible size.[31]

Inequalities of wealth and power here, for both authors, generate significant negative effects, forcing "all ethical defences of freedom [to] lose their validity", and allowing "political liberty … to lose its value". There is considerable intensity in both authors'

[30] Rawls 1971, 287.
[31] Knight 1935, 291. The phrase "a cumulative growth of inequality" is underscored in red in Rawls' copy. (David Levy Scans)

critiques. This intensity goes largely unremarked upon in analyses of Rawls' system. Rawls' final diagnosis of the political system is entirely Knight's. The next passages cited are just after the passage quoted above. They are not sanguine.

> Historically one of the main defects of constitutional government has been the failure to insure the fair value of political liberty. The necessary corrective steps have not been taken, indeed, they never seem to have been seriously entertained. Disparities in the distribution of property and wealth that far exceed what is compatible with political equality have generally been tolerated by the legal system. *(*Rawls 1971*,* 226)

Rawls then restates Knight's position, comparing political processes unfavorably to those of the market ("Essentially the fault lies in the fact that the democratic political process is at best regulated rivalry; it does not even in theory have the desirable properties that price theory ascribes to truly competitive markets").[32] Fortunate historical periods of equality will be quickly undermined. Universal suffrage is "an insufficient counterpoise".[33] "Basic measures needed to establish just constitutional rule are seldom properly presented" because "the political forum is so constrained by the wishes of the dominant interests". Politics so characterized is obviously not an argument for possibilities outside the original position. Whether these inclinations are termed "interests" or "seeking after power" (and for Knight these are identical), Rawls views the inevitability of their influence as strong

[32] Rawls 1971, 226.
[33] Identically in Knight: "Equal suffrage" provides "little or no guarantee of equality…." Knight 1935*,* 291. ("Equal suffrage" underlined in red twice by Rawls, in the middle of very heavily marked pages) (David Levy Scans)

indications of the need for isolating the original position.[34]

IV. Discussion, Consensus, and the Original Position

This section, centering on decisions made in the original position, has a number of involved and complex strands, so will be simplified in the following way. I will focus on two facets of decision: a) belief in a goal to be reached, and b) consensus vs. simple agreement. These two dimensions in Rawls and Knight point to the philosophical as well as economic overlapping of their arguments.

a.) Beginning with Knight, we see an effort to ensure that concepts like "objectivity" and "truth" aren't partitioned away from social questions, to be located solely in the domain of science. Knight rejects attributing objectivity only to science. For him, investigations of scientific and social questions are linked.[35] Although there are valid discriminations between them in terms of reliance on data, and objective testing, they have dramatic and critical similarities. Both investigations rely on "values", and both achieve validation through some form of consensus. Knight was deeply read, and he was

[34] In his lectures on Locke, Rawls makes the assertion that Locke's social compact is marred by exactly this problem: individuals retain bargaining advantages within the compact-forming environment, resulting in a post-compact class structure. See the three lectures on Locke (Rawls 2007), particularly 151-2 and 155.

[35] We might be curious about the level of reliance Rawls places on Knight. Fleischacker provides one reason: "When Rawls started writing, pretty much only Marxists and utilitarians were willing to develop normative accounts of political issues, and even they were under constant siege by the upholders of the reigning positivist paradigm, for whom all normative declarations were expressions of emotion and did not belong in scientific or philosophical analysis." (Fleischacker, p. 110) And on the question of relation to positivist doctrines: "Thus, whereas the logical positivist of thirty years ago would insist that the economist *qua* scientist must avoid value judgments, on the grounds that such judgments cannot be scientifically justified, Rawls would argue that value judgments verify in much the same way as do factual hypotheses." (Worland, 122)

skeptical of science as an "absolute" description of reality – his phrase was to describe scientific (and social) theories as "relatively absolute absolutes." Knight's analysis, in fact, dissolves everyday realities:

> The attempt of science to find what is real in human behavior reduces it first to mechanical movements and physiological processes, in themselves sufficiently different from the "immediate" experience or observation of life. The rest is inference and emotion. But physiology just as inexorably dissolves into chemistry, and chemistry into physics, and all that physics leave of reality is electric charges moving in fields of force – things far more unreal than the characters in the most fanciful works of fiction. Moreover, the experts in science and scientific method (vide Mach, Pearson, Russell) are frankly skeptical of the reality of any of it, and talk in terms of concepts useful for the purposes of analysis, and of the simplification of our thought processes. (Knight 1925, 396 in the *QJE*, 94 in the *Ethics* collection)

Science was for him less a discovery of bedrock reality than it was a mode of thinking, a development of a means of "analysis" (94). It is, nonetheless, capable of reaching conclusions. This is true for social questions as well. Knight explicitly attacks the dichotomy:

> In view of the virtual deification of science, in modern thought, as the only mode of valid intellectual activity, the point needing emphasis is the large number of kinds of mental activity which have to be regarded as intellectual and affected with validity. The black-and-white dualism of the modern empirical-utilitarian

> world view – the notion that every statement relates either to a physical world in which truth is absolute or to "subjective" preferences, any ascription of validity to which is either illusion or arrogant presumption – is a major heresy of our civilization. The truth is rather that opinions in both fields have greater or lesser degrees of validity. Truth is an ideal in which we must believe to give meaning to thought and to life; but there is no way of knowing that any particular belief is true, and every belief must be held subject to revision – except the belief that there are better and worse reasons for believing. (Knight 1935, 346-7) (This quote heavily underlined and margin-marked in red ink by Rawls.)

So for Knight the same standards of analysis, and the same hope of reaching conclusions, span both the technical sciences and social (and moral) investigations. Given common presuppositions, this has the dual effect of rendering science less "scientific" (in the traditional sense) than before, while social and moral questions become more scientific. In the terminology of modern analysis, these latter questions of social and political morality become more *tractable.* And Knight includes economics, viewed from a certain vantage, as a discipline partaking of this more complex involvement with meaning and "values". The "science" of economics he considers mere mechanics: it has use as a standard, but human actors within its framework aren't really human – he calls them pinball machines! But economics more fully considered does contain value and meaning. For instance, economic activity is a matter not merely of selecting among pre-existing wants, but must instead involve both desires and causes of desires.

> They [wants] have to be thought of and treated as much more than forces, conscious or unconscious, which dissolve into mere phenomenal uniformity of coexistence and sequence. On the one hand, desires have a primary, assertive, creative, and experimental character; they are choices. On the other, they have a cognitive quality. (Knight 1925, 399-400 in the *QJE,* 97 in the *Ethics* collection)

Skepticism is as correctly applied to scientific as to social-value conclusions, but we cannot be complete skeptics and live. Knight's "relatively absolute absolutes" is present in all searches. In both science and social science, as well as straight morality, the goal is conditioned by, and directed towards, values. For Knight our deepest "wants" are those we create, and the process of want-creation is at the base of the transition of economics from a simple mechanical description of "forces", to a description imbued with human significance. This for Knight also forms the weakness of what he terms the empirical-utilitarian view. This view sees choice as between baskets of "goods". When those goods are envisioned as being "created", as being more process than result, then the choice-calculus begins to break down. Knight campaigns against "the assumption that human wants are objective and measurable magnitudes and that the satisfaction of such wants is the essence and criterion of values."[36] Utilitarianism for Knight, like economics in its "science" configuration, is fundamentally mechanical in nature. Ethical and value dimensions are excluded almost by definition. We are left with the ability to calculate, but such calculations can only be completed in a world of "given" wants and ends. There is a leap beyond pure calculation that must be made. It is such a leap, made by individuals

[36] Knight 1923*,* 579 in the *QJE,* 41 in the *Ethics* collection.

behind the veil of ignorance in Rawls' scheme, that causes utilitarian and Bayesian critiques to miss their mark.[37]

Much of this would clearly be in step with Rawls's inclinations, and the direction of his project. Specifically, beginning with the last point, Rawls endorses the complication that want-creation represents:

> Moreover, the social system shapes the wants and aspirations that its citizens come to have. It determines in part the sort of persons they want to be as well as the sort of persons they are. Thus an economic system is not only an institutional device for satisfying existing wants and needs but a way of creating and fashioning wants in the future.[38]

These sentences occur in a section entitled "The Concept of Justice in Political Economy". It could, however, easily be descriptive of the function of choice in the original position. What Rawls intends us to do under the circumstances it presents is not merely import conclusions about principles for social structuring, unmodified, that we already hold. He supposes, rather, that some views will partially conflict with others, as they must. We are then to let more deeply held notions interact with those less deeply held, and so modify and alter them. The end of result of this reflective deliberation is a

---

[37] It also highlights what is, and isn't being appropriately decided in the original position. Debates about particular post-original position gains or losses are deprived of their calculation basis. For debates still concerned with winners and losers, see the following. Mueller (1989, p. 417) argues that the sacrifice that causes problems is the one the rich are asked to make to benefit the poor. This notion is also mentioned in Nagel (1974) and Scanlon (1974). Knight's power arguments, or his leap from calculation to value, are not mentioned. A paper more fully exploring this connection is in process.

[38] Rawls 1971, 259. A further quote in Knight was exceptionally heavily marked by Rawls: "The development of wants is really much more important than their satisfaction: there is no poverty so deplorable as poverty of interests." (Knight 1925, 407 in the *QJE,* 103 in the *Ethics* collection)

new synthesis, an equilibrium.[39] Knight has a parallel argument about beliefs and consciousness: we operate in our network of beliefs, for the most part, fairly unconsciously. These beliefs might contain inconsistencies, or be incompatible with other beliefs, but these problems are not troubling because we are operating largely unconsciously. But once we focus consciously on a belief, we can no longer retreat to our unconscious acceptance. We must now work out a more "deliberate and rational" justification.[40] Something closely paralleling this is happening with Rawls and the idea of reflective equilibrium.

As for "truth", or the concept of being able to reach best outcomes, Rawls argues that the first step is the "fairness" in "justice as fairness". By this he means the principles of justice that would be chosen from inside a fair choice environment. He must first justify the original position:

> The concept of the original position, as I shall refer to it, is that of the most philosophically favored interpretation of this initial choice situation for the purposes of a theory of justice. (Rawls 1971, 18)

---

[39] "In describing our sense of justice an allowance must be made for the likelihood that considered judgments are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even through the theory does not fit his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral philosophy, the best account of a person's sense of justice is not the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium." (Rawls 1971, 48)

[40] Knight 1935, 347.

Much follows from how the initial position is characterized. In fact, each of the systems that compete with Rawls' own would have its distinctive attributes, which would be reflected in different original positions. So Rawls can affirm that, given the variety of starting points, his system is but one of many. This is also what dramatically separates Rawls' system from those which do not describe an initial choice environment. The principles chosen behind the veil are "the only choice consistent with the full description of the original position."[41] So, in a straightforward way, everything the system concludes hinges on how the original position is specified. Particular principles are then reflective of that original environment. Yet this is still essentially predicated on their being something beyond the "purely subjective" and wayward variety of principles which might be arrived at, no matter how the original position is restricted. Justification for acceptable "conclusions" being out there at all is substantiated by, among other inputs, arguments from Knight. It was from Knight, after all (according to a statement from Rawls), that the deliberative choice environment was first envisioned.[42]

b.) In the "political" section, the difficulties that plague open discussion of principles were briefly sketched. Knight is wary of "persuasion"[43] or bargaining as means of reaching principled conclusions; these are elements of a contest, where power and not principle is the relevant currency. Knight's answer to this problem is to attempt to refine the nature or environment of the choice: it should be made by disinterested actors, who are specialists in such deliberations. Rawls answers these problems by also characterizing

[41] Rawls 1971, 121.
[42] See footnote 5.
[43] Knight 1935, 345.

the conditions of choice – this is his original position. And because the veil screens off the particular interests of participants there, in the original position we escape distortions from bargaining and power imbalances.

> Thus there follows the very important consequence that the parties have no basis for bargaining in the usual sense. No one knows his situation in society nor his natural assets, and therefore no one is in a position to tailor principles to his advantage. (Rawls 1971, 139)

Discussion, therefore, although in an ideal sense being what one wants, is ruled as inadvisable in a real-world environment. And because agreement could be reached through means of discussion, agreement itself as a criterion of excellence is under suspicion. Discussion can take place on any number of different "levels"; it is Knight's and Rawls' contention that the least desirable levels might be the operative ones. Knight guards against this through restrictions about which he doesn't seem particularly optimistic. Rawls has instituted much more extreme precautions. Rawls feels that if he can successfully restrict consideration to the appropriate contemplation of principles, then suitable outcomes could be anticipated.[44] Rawls, however, faces very different challenges than does Knight. His system is presented as being in the tradition of contract; it would seem at first blush that ruling out ordinary discussion would put a dagger through the

[44] Though is it possible that too many restrictions have been imposed? Barber argues for this view: "Now there is a considerable question in my mind about whether it is possible to conceive of men as having a hypothetical knowledge of what it means to have interest and desires without having particular interests and particular desires. Mutually disinterested men might turn out to be uninterested men, men incapable of comprehending the meaning of interest. Rawls suggests as much when he concedes that 'some may object that the exclusion of nearly all particular information makes it difficult to grasp what is meant by the original position' (138). At the level of psychology it seems possible that particularity is built into the notion of interest and that it cannot be cut away without rendering interest unintelligible." (Barber 1974)

entire enterprise. Yet Rawls follows Knight's progression in an odd way: Rawls restricts the "variety" of his individuals' rational deliberations by stripping away those factors which would generate differences. Just as Knight restricts the numbers involved in his deliberations to achieve a greater consensus (and a "higher quality" consensus) of opinion, so Rawls whittles down his innumerable individuals to a single deliberator. The speed with which this is argued is worth reexamining:

> To begin with, it is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. Therefore, we can view the choice in the original position from the stand point of one person selected at random. (Rawls 1971, 139)

In just a few sentences, Rawls achieves the unified viewpoint that Knight struggles over for an entire essay. The beauty of the original position, with its veil of ignorance, is that the barriers to consensus are ruled extraneous by definition. It is not the purpose of this paper to argue whether these restrictions are excessive, or even feasible. But they are an intriguing method for gaining consensus so complete it can be termed unanimity.

Knight struggles to get his decision-situation configured so that principled conclusions are possible. His conclusion posits an elitist retreat to a restricted set of "experts". Rawls, as a contractarian, cannot take this path. But he finesses the difficulty: each person is expert enough, if we pare away distracting focuses on personal interests, and situate the principles themselves within the reach of every person's common sense. The generality of this decision is further accentuated through the notion of publicity: each idea must be acceptable to all, and the individual in the decision-environment must take

that demand into account. In this manner, the fact of agreement behind the veil is transformed into a concept of consensus; we now are not in the realm of power conflicts, but in the realm of value. Knight's concerns have been addressed; Rawls has defused the conflicts real-world interactions would engender – the dimensions of power – by creating a decision-field that removes the knowledge that would fuel rivalry and influence. For Rawls in a hypothetical contractarian experiment, agreement would be a difficult enough hurdle; to stipulate further that agreement itself carries no principled significance would seem to make the task insuperable. These are some of the issues Rawls highlights in *Ethics*. Yet Knight's half-hearted restrictions point the way to Rawls' own much more extensive maneuvering. By defeating the interest-conflict sketched by Knight, Rawls also manages to unify the rational deliberations of his actors behind the veil. In one fell swoop he achieves unanimity while preserving the fragile contractarian basis of his system. Whether this might be considered a coup too far is certainly debatable. No matter what the assessment of these innovations, however, the intersection with the arguments and concerns of Knight's makes Rawls' strategies more readily apparent.

V. Conclusion

Rawls' original position is one of the most striking, and central, features of his system. The functioning of his theory is predicated on it; choices between principles which might be clouded in real-world settings become clarified behind the veil as he envisions it. Without knowledge of social position or particular psychological dispositions, reasonable deliberation in a state of "reflective equilibrium" can lead to acceptance of the

principles he proposes. Rawls does not claim that this is the only possible view of justice that might prove acceptable, but it is one of them. Appraisal of the original position concept frequently begins right there, with questions about how reasonable it is as a psychological state, or how its premises could generate completely divergent conclusions. This paper has focused, instead, on looking at the alternatives to the original position, how the original position addresses shortcomings Rawls discovers in these alternatives, and how these arguments tie together with similar arguments in Knight. These were broken down into three main areas.

First, when discussing the market, Rawls makes the unusual choice of faulting market outcomes, but preserving faith in markets as a process and benchmark. He also uses markets as his example, in the efficiency half of the efficiency-values continuum. But efficiency virtues of markets break down when transitioning from allocation to distribution. Too many factors – from initial wealth allotments to luck to supply and demand or scarcity being unrelated to deservingness – are present to permit any reasonable link between market outcomes and moral worth. This is a combination of both exterior inputs being unavoidable, and real-world markets not fully preserving the virtues of their ideal counterparts. Rawls examines and presents these arguments in considerable detail. He concludes that the failure of the market to provide ethical justification for its distributional outcomes necessitates conditioning decisions made previous to such social structures, decisions needing to be made in an atmosphere characterized by reflective equilibrium, behind the veil.

Secondly, the political arena is considered as a setting for ethical deliberation. Rawls in *Justice* is discussing the political dimension "in turn"; that is, he examines the various stages of deliberation, and the political process is one of them. But, implicitly, the reader is guided to assess whether a "later" stage might possess the structural conditions necessary to make it a particularly promising stage for ethical deliberation -- or not. And if this proved to be the case, the question would arise of whether it might replace a more problematic, less realistic "earlier" stage. While political institutions might seem a promising alternative, it turns out that political decisions share the faults of market participants' striving for power, but to a still greater extent. The urge for power is even more focused than it is in the market. Its failure is a strong argument for considerations of justice needing to be built in before this stage is reached.

Thirdly, discussion as an ideal is undercut by a) interests, and b) the tendency to reach agreement through bargaining. Bargaining (as a trading of interests) and agreement per se are not processes assured of ethically compelling outcomes -- quite the reverse. The ideal of the single deliberator is achieved through the veil, where a lack of interests and knowledge of social position assures that all actors can be modeled as a single actor, and a form of unanimity result. Reliance on common sense and reflective equilibrium implies that sufficiently "expert" judgments can be realized, without the elitist connotations that concept implies. Individuals, then, are empowered in some sense through isolation; in the environment where social positions are known, and others present for discussion and bargaining, nothing of ethical consequence, Rawls believes, could be accomplished.

Achieving the specifics of this helpful isolation is the characterizing of the original position.

It has been the purpose of this paper to elaborate these ideas from a particular perspective: through the lens of certain essays by Frank Knight. Through matching of quotes and noting parallel lines of argument, it tries to show Rawls' indebtedness to Knight. And these parallels have been able to be rendered even more emphatic through Rawls' extensive annotations of Knight's essay collection. But beyond being a source, or crucial reinforcing, for ideas in *Justice*, it is hoped that the perspective through Knight has set out some of Rawls's ideas in greater relief. This paper has focused on the arguments against alternatives to the original position – there could have been other interesting focuses. Rawls and Knight share an interest in using the idea of the "fair game" as a metaphor. They each discuss "want-creation" as opposed to "given" wants. The role of ideal types in Knight, and ideal procedures generally, is often highlighted in Rawls's text. The role of such ideal types, and how Knight emphasizes their isolation, relates, perhaps quite directly, to Rawls's original position. For Knight the economic actor is "not a social animal", and is thereby isolated from/immune to feelings of rivalry or emulation.[45] This sort of immunity is crucial in Rawls's deliberative procedure. The isolation Knight postulates therefore ties closely to the "independent deliberation" Pogge describes as one of Rawls's original formulations (and which Pogge indicates Rawls acknowledged came from Knight).[46] Each of these presents interesting potential paths for investigation.

[45] Knight 1935, 282 and 295-6.
[46] For useful discussion on Knight's transition to using Weber's ideal types in his "Economic Theory and Nationalism" I am indebted to Ross Emmett.

Rawls's justifications for his theoretical constructs, usually elaborated extensively in his text, are necessarily summarized by his critics. It is unfortunate that these summaries tend to skirt the supporting arguments from economics, but they fail most particularly in failing to consider Knight's idiosyncratic views. A reading of *Justice* with them in mind makes their exclusion seem almost a distortion of intent. In any case, their inclusion renders *Justice* a clearer and more coherent experience.

2. MODELING IN RAWLS: THE ORIGINAL POSITION

I. Introduction

Who are the individuals in Rawls's "original position"? A reader's answer to this question frequently determines a judgement 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" (*Theory,* 19). 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. The difficulty for the reader is that 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 intention), 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. But 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.

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.* 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, 162-3)

It is important to realize that Rawls's original model for his investigation — 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, 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, vol. IV, 327)

Rawls, looking for a path into his problem, was certainly aware of these simplifications and abstractions of Mill's. It 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![47]

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

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

> 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 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, 164-5)

This tendency towards abstraction would only become more extreme in the contemporary economists with whom Rawls was also familiar (for instance, in Buchanan, Arrow, Sen, Musgrave, Baumol, Meade, etc.). 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 more particulars of the argument, it might be useful to briefly touch on Rawls's use of economics as a reference point. Whenever such references surface in the work, it is surprising to his philosophical readers, but also to his economic ones. Such references, no matter Rawls's reliance on them, tend to be excluded from how critics summarize his arguments. For instance, when Rawls is considered 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. (Rawls, 1971, 308)

For his analysis of politics, the starting point is again economics,

> 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. (Rawls 1971, 359)

By the next page we will find out that politics, of course, is *not* a perfect procedure.

Discussing the original position, again an initial comparison is with economics,

> I now turn to matter 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 common basic technique in economics. He will go on to discuss the ability to predict the economic actor's choices (prediction will be a theme later in this essay). 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." 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.… (Rawls 2001, 82-3)

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 predict 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 should be judged by 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.

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" (2001, 17). 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…. (2001, 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. (2001, 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,

> 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 brief that equilibrium is necessarily more important than disequilibrium, but rather that the analysis of behavior in disequilibrium is substantially more difficult. (Varian, 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.[48] 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 ascendency 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, 13-14). And a number of economists in the 1950's and 60's were beginning to address questions that interested Rawls, bringing with them a mode of thinking closely linked to their training in and exposure to 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*,[49] Sandel 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."

> 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

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

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

> 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, 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 (*Theory,* 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 judgements. They must also be acceptable to others. But this process is to modify the original position itself; if we cannot reach "significant" judgements, 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, 48-9)

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 can't 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" (Sandel, 62). 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.

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. This predictive side of things also clearly ties to economics. There is a dominant strain in modern economics that seeks a firm connection between premises and outcomes, and focuses on prediction. For practitioners of a would-be science, the attraction of the latter is obvious. Milton Friedman famously suggested that in fact prediction was all that mattered; what assumptions got you there were largely irrelevant (Friedman 1953). This ability to predict is often seen as the gold standard of what constitutes "real science". In philosophy of science, the centrality of prediction is usually traced back to logical positivism and the Vienna Circle. Caldwell summarizes their position,

> The logical positivist program asserted that only meaningful statements were to be permitted scientific consideration and accorded the status of knowledge claims. Meaningfulness (or cognitive significance) was strictly defined as being attributable only to those statements which are either analytic (tautologies or self-contradictions) or synthetic (factual statements which may be verified or falsified by evidence). By this criterion, metaphysical statements are neither analytic nor subject to empirical test, so must be deemed meaningless, expressing emotional stances or "general attitudes towards life" . Caldwell, 13) (the quoted phrase is attributed to Carnap)

And the logical positivists have a direct connection to Rawls. In a ground-breaking essay on Rawls's graduate school papers, David Reidy isolates Rawls's use of the ideas of the Vienna Circle.

> Rawls characterizes his concern with and ambition for ethics as science as fully scientific in the sense associated with the Vienna Circle. That ethics as science unfolds within the space of reasons rather than causes is irrelevant to its status as science. What is crucial, Rawls insists, is that it must avoid all theoretical claims that can neither be confirmed nor refuted by publicly observable evidence, in particular the evidence constituted by familiar, everyday, noncontroversially competent moral judgments. (Reidy, 14)

In these early efforts, Rawls observes (according to Reidy) moral philosophers, to see what their procedures are. Those procedures appear to do two primary things: they are 1) "engaged in a science of moral judgment", and 2) "they seek to explain competent moral judgments in a way that would enable us reliably to predict them" (Reidy, 13). The procedures are *not* engaged with a host of questions that concern analytical philosophy, those "associated with Bertrand Russell, G.E. Moore and the early Wittgenstein." The sorts of models (Reidy uses the word models) that *should* interest moral philosophers could be characterized as "reasoning machines" (Rawls's phrase). In this phase of his thinking (1946), Rawls is not concerned with the *justification* of either the principles resulting from the "reasoning machine", nor the justification of the reasoning machine itself. The "data" which it is supposed to predict consists solely of *noncontroversial* moral judgments. The machine, then, simply generates principles which can distinguish between *competent* moral judgments, and *incompetent* ones. A useful reasoning machine will output principles that can make this discrimination.

Rawls's use of the phrase "reasoning machine" for this early version 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 present in *Theory* as well. A somewhat lengthy quote covers a number of 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 assumptions 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. (1971, 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 as a second best approximation of his actual model: moral geometry. But we should also remember how psychological assumptions were mentioned when Rawls's discussed economics. In that context they implied no attempt at verisimilitude; nor should they have such 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 doesn't mean they can't have real-world implications, however. Geometry, for instance, can aid in such things as bridge building (Hands, 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" (1971, 121)). Rawls has *constructed* the one which will output his desired principles.

To distinguish between these initial situations, between the various ways in which they might be specified, is the work of what Rawls terms (in early work) "ethical theory" or "moral theory" (Reidy, 18), which can be understood as a next stage after "ethics as science". But if the nature of models (for Rawls, ethics as science) is often so contrary to intuition, and predicts what we already recognize and accept, why not simply do without them? But the obvious predictions are a test of the machine; it will have a function when we come to controversial judgements. Testing, however, must come first. Rawls explains this with the analogy of linguistics. As native language speakers, we possess a sense of

grammar from habits of use, and from what we hear and read. Difficult questions are, however, often unresolvable on these grounds alone — different individuals may simply have a different sense of what sounds or seems correct. Linguistic rules are required to sort out some of these issues. Rawls argues that moral philosophy is similarly situated. As in his system, linguistics needs rules that can "make the same discriminations as the native speaker. This is a difficult undertaking which although still unfinished, is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge" (*Theory,* 47). Thus the elaborate construction, and abstraction, of the original position are unavoidable.

> 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. This is to be expected, since on the contract view the theory of justice is part of the theory of rational choice. (*Theory,* 47)

The straightforwardly technical character of the foundation of Rawls's thought is here difficult to overlook. And of course contentious issues can result from parties arguing purely self-servingly, or from a bias concerning race or group. Rawls intends his theory to clarify these. And as Reidy suggests, this emphasis on rationality has been present in Rawls from his earliest work, "*our moral nature is part of our rational nature,* not something that unfolds within the space of facts, events and causes" (Reidy, 20-21,

emphasis added). Morality here is subsumed within rationality. This rationality represents an expanded version of self-interest; self-interest linked to notions of generality, reciprocity, and publicity. (Later, for Rawls, this will be the rational subsumed within the reasonable.) This primacy of rationality creates another strong link to economics.[50] Adam Smith's view of self-interest also is situated within a larger structure of reciprocity. And both Rawls and Smith explicitly reject benevolence as an operational principle.

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. Rawls's reasoning machine is clearly demarcated as such to avoid criticism of its (not *his or her*) failure to resemble actual persons. It is constructed to generate output of a certain sort.[51] Here we can find echoes in Friedman's methodology paper, as pointed out above. Friedman's paper was widely influential, and its position would have been implicit, if not explicit, in many readings Rawls encountered. Its emphasis on prediction as the only meaningful evaluative procedure is the dominant message. Friedman of course goes further than most expositors of positivism, in so downgrading the importance of realistic assumptions as to suggest "the more unrealistic the better", as long as the predictions pan out. Friedman's "overstatement" was not widely endorsed, but his essential point about predictions was.

[50] "We suppose that the parties are rational, where rationality (as distinguished from reasonableness) is understood in the way familiar from economics." (Rawls 2001, 87)
[51] "We want to define the original position so that we get the desired solution." (Rawls 1971, 141)

Rawls must have found such methodological positions congenial when his serious study of economics began.[52]

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)

What Nagel here sees as a too individualistic foundation to the original position is simply a blind alley. He makes the critical mistake 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

---

[52] Friedman's position that assumptions don't matter serves to break the causal connection back to actual motivation. His science is *forward justified* — by prediction. The argument is not justified by going behind the model, to find and establish the causal roots. Rawls endorses this view, and further from his economic roots, terms it a Kantian one in *Political Liberalism,* "The absence of an explanation in cognitive psychology is not to the point: being able to give the proof, or to state sufficient reasons for judgment, is already the best possible explanation of the beliefs of those who are reasonable and rational. At least for political purposes, there is no need to go beyond it to a better one, or behind it to a deeper one." (Rawls 2005, 120)

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 (Freeman, 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 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 does (1998). 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". Similarly, O'Neill recognizes the "no actual persons take part in or count in the Original Position" (Nussbaum, 5 and 17; O'Neill, 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 realization, of the mechanics of the original position, should divert our critical attention elsewhere. Whether we are comfortable with it or not, it forces us to be outside the system, 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 is 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. The predictive aspect, however, should reassure. That is, when we judge noncontroversial moral situations using the outputted principles, we approve the result. They have been vetted in that important sense. We then move to other questions: 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 by picking at the verisimilitude of a process when that verisimilitude is beside the point.

A somewhat unusual parallel might prove informative. Morgan in her models book devotes a section to models as caricatures.[53] 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.

[53] 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."

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.[54]

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 person. It models, you could say, a single dimension of its subject — his fatheaded-ness. 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. Our imaginary playmates are every bit the fleshed out friends we want, and imagined scenarios 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 determines how the "person" winds up in the model, not the strength of the connection back to real individuals. If economic actors

[54] Morgan, 157-164.

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, in reality, 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)

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.[55] Unanimity is a very mechanistic result. 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 *Theory*, and with the authors of which he initiated a correspondence.[56] 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.

---

[55] "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." (Rawls 1971, 141-2)

[56] 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).

But how to achieve the virtues of unanimity, absent the costs? One technique which probably 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 *Theory.* 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.[57] 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 seeks and these everyday muddying concerns would seem to be intertwined. Modeling is Rawls's way to cut the Gordian Knot.

V.

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

[57] I'm using the word "noncontroversial", from Rawls's early work, as a shorthand term. In *Theory* 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…." (1971, 19)

to the actual world to be of interest, but their connection does not appear to be exactly the same. Morgan sees these as distinct (Morgan, 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 lot" of buyers were price takers. If some actually 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; 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 a competing — a different — theory of justice. By not selecting it, when he could have, Rawls implies this alternate model would prove less robust than his own. We must suppose he considered all of 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 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" (Morgan, 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" Sugden 2002, 117). For Rawls, this would seem to imply looking to applying principles to the noncontroversial moral situations. Or, as Rawls phrases it in *Theory*, the output must align with certain convictions. And, "these convictions are provisional fixed points which we presume any conception of justice must fit" (1971, 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 are immune to 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. (Rawls 2005, 27, emphasis added)

Rawls then introduces the comparison with role playing. When acting the role of Macbeth or Lady Macbeth, we shouldn't be thinking we actually 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" (Rawls 2005, 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 "state of nature" (1651) 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 probably more extreme than most.

3. INFORMATIONAL RESTRICTIONS IN RAWLS'S ORIGINAL POSITION: ECONOMICS AND THE INCONSISTENT PLANS LITERATURE

I. Introduction

The original position in Rawls's *A Theory of Justice* (henceforth, *Theory*) appears at first to be a more extreme assumption than is necessary to formulate Rawls's two principles. Surely, in possession of all our personal histories and views, we should be as able, or more able, to come to terms with such questions. How can blocking off portions of experience possibly improve our ability to gain perspective on deep social concerns? One way to see Rawls's strategy in this instance might be to look at similar strategic assumptions, in other parts of the social sciences. That is what this paper will investigate. The inconsistent plans literature, where a similar informational restriction (distancing) is assumed to improve decision-making, serves to support Rawls's assumptions. In both cases there is a problem to solved. And the solution in each case involves a restriction on the decision environment. And in Rawls such informational restrictions then shape self-interested rationality, to determine decisions behind the veil of ignorance.

But there is a question preliminary to this. Why is there such difficulty not only in coming to terms with Rawls's argument, but even with determining what exactly that argument is? There is a particular challenge reading *A Theory of Justice* as a text. Rawls employs various descriptions to explain his concepts — goes down different expository routes, as it were. This serves to make the text more available to a diverse set of readers.

The difficulty is that these differing paths of explanation sometimes have differing implications for the theory itself. Descriptive strategies lead to adopting overviews which Rawls assumes we can dispense with when necessary, when rendering final verdicts. These various descriptive pathways should assist us, not trap us. But such pockets of interpretation remain open in the text. Rawls will summarize a way of conceptualizing something, emphasize its importance, and then ignore it in subsequent passages on the same topic. Rawls himself offers a dual perspective. He utilizes particular explanations widely in the text, but then tells us to think about them in a quite different way. An additional issue is that these summaries are often based on the economic side of his analysis, placing them farther from the core competency of the majority of his readers. Yet ignoring them is fraught. This paper argues that Rawls utilizes economic modeling as the template for his own system. This is a substantial claim, but Rawls offers considerable evidence for it in his texts. Reading through the economic lens, like any choice of perspective, requires almost a literary attentiveness, as well as analytical. As in a novel, elements that are dispersed can come together meaningfully. But such interpretive efforts invite missteps. The test will be whether the whole satisfies, and enlightens. We can only hope that, as Rawls might say, the more we consider it, the more correct it will seem.

The immediate topic of this paper — informational constraints in the original position — is embedded within this overall economic framework. The paper places the inconsistent plans literature — a literature spanning economics, psychology, and philosophy — against Rawls's own "planning" dimensions, and compares them. The

literature in question largely post-dates Rawls's *A Theory of Justice*, and in any event, he does not footnote the papers he might have. But his own discussions of the various topics — time-discounting, plan consistency, multiple selves — are remarkably in step with the papers which appeared after his publication of *Theory* in 1971. I think we can learn something from these similarities. The concept of rationality, for one, plays a central role across all these texts. And in particular we will see that the strategy used to formulate successful plans — a distancing from immediate circumstance, and a more abstracted, long-run view of choice — provides restrictions on information that serve much the same purpose across these authors. The informational restrictions provide the environment within which superior decisions — in both this literature and Rawls — can best be made.

Not all elements of economics conform with Rawls, or with the inconsistent plans studies. For instance, standard teaching in economics has agents, attempting to maximize their utility, benefiting from operating in a full information environment. And as before, we might ask: how could *not* knowing something improve decision-making? Yet Rawls, and our inconsistent plans authors, make exactly this argument.[58] There individuals are screened from certain informational dimensions, in order to *enhance* the quality of decisions. One could frame this stance as an argument against locality — it is of course not information *per se* which is obstructive, but information of a certain type. Two sorts of information are to be considered. The first is located in the everyday world, a world of "momentary stimuli" (Schelling, in "The Mind as a Consuming Organ"). The second

[58] Rawls is explicit about the problems of full information when unanimity is the goal (as it is in his theory). "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." (*Theory* 141; *TJR.* 122)

consists of more abstract or general information, usually contemplated from some place removed, out of the fray. For inconsistent plans studies, the abstraction is due both to physical distance (no locality), and to separation in time. These two serve as a filtering process. Rawls will create his abstraction with a more extensive filter. But the processes are similar. Decisions from this filtered environment are assumed wiser, or more correct. It is helpful to remember the words filtration, or distillation, rather than restriction or screen. This helps us see through what at first appears its paradoxical nature. It is the exclusion of immediate stimuli that permits a more abstract view, and which leads to the decision's greater potential generality. On the other hand, immersion in situational specifics — in locality — pulls towards personal interests and particular concerns, and away from general ones. At least this is the assumption shared by Rawls and the majority of the studies we will examine (but not all). Certainly Rawls aims for generality in his theory. The question will be, whether the approaches analyzed in these studies of the divided self parallel his efforts sufficiently to cast illumination. The argument will be that they do.

The tie to studies in other disciplines — primarily economics and psychology — might at first seem too remote from Rawls's philosophical project. Yet Rawls himself continually connects his arguments with those from other disciplines, particularly economics. Such bridges are so numerous that one begins to see them less as supportive, and more as foundational.[59] For instance, this paper focuses on the original position. In

[59] Rawls discusses Frege and Cantor, and their advances in logic and set theory, as models for where analysis of "moral conceptions" might eventually go. Despite his admission that study of moral conceptions is currently "primitive", still, the idea of the potential goal for analysis is

his chapter on the original position in *Theory*, Rawls, after emphasizing a little-remarked-on dimension of that position,[60] begins one paragraph,

> By arguing in this way one follows a procedure familiar in social theory. That is, a simplified situation is described in which rational individuals with certain ends and related to each other in certain ways are to choose among various courses of action in view of their knowledge of the circumstances. (*Theory* 119; *TJR* 103)

But where in social theory are such "familiar" procedures? The answer comes two sentences further on:

> In the theory of price, for example, the equilibrium of competitive markets is thought of as arising when many individuals each advancing his own interests give way to each other what they can best part with in return for what they most desire. Equilibrium is the result of agreements freely struck between willing traders. (*Theory* 119; *TJR* 103)

One sees in this short passage considerations that will be central: self-interest, a "simplified situation", rationality, and equilibrium. Rawls proceeds to enumerate other characteristics of markets that are of use: ideal markets respect the "right and freedom" of others. Equilibrium there follows because individuals have reached a "best situation". In fact, his own use of these elements faces additional problems. For instance, equilibrium, in social terms, could simply result from "a balance of hatred and hostility", and stability

---

highly abstract and logical. We think of Hume's and Adam Smith's more hard-edged theorizing, with Newton as at least partial inspiration. (*Theory* 51; *TJR* 45); see Schliesser (2020) for a summary of Hume and Newton)

[60] The idea is that the reasoning taking place in the original position is "strictly deductive" (*Theory* 119; *TJR* 103); that is, that no actual deciding in the common sense of that word takes place.

of this sort would fall outside the goals of Rawls's project. His principles must be "acceptable from a moral point of view" (*Theory,* 120; *TJR* 104).

Exploring this economic avenue for a moment longer brings us to a surprising juncture. When introducing his concept of rationality, Rawls uses similar phrasing to the above; to wit, the idea of rationality he draws upon is "the standard one familiar in social theory" (*Theory* 143; *TJR* 123-4).[61] Again this means primarily economics.[62] The footnotes for this are to the economists Amartya Sen and Kenneth Arrow (1963). Looking up the Sen, however, on the pages indicated we find this:

> … it will be a mistake to assume that preferences as they actually are do not involve any concern for others. The society in which a person lives, the class to which he belongs, the relation that he has with the social and economic structure of the community, are relevant to a person's choice and not merely because they affect the nature of his personal interests but also because they influence his value system including his notion of 'due' concern for other members of society. The insular economic man pursuing his self-interest to the exclusion of all other considerations may represent an assumption that pervades much of traditional economics, but it is not a particularly useful model for understanding problems of social choice. (Sen 2017, 50-51)

---

[61] By *Political Liberalism* Rawls will replace his idea of rationality with the combination of the rational and the reasonable. The updated formulation revolves around the idea of an overlapping consensus. This has a less mechanical feel to it, certainly, but its fundamental difference from his earlier configuration in *Theory* does not strike all critics as a huge displacement of his original concepts. I tend to agree with this assessment.

[62] "Economic theory is both the birthplace and the prime application of the rational choice paradigm. Throughout the 20th century, economics has relied on rationality, and refined the definition of rational choice…." (Gilboa, Postlewaite, and Schmeidler 12)

This strikes us at first as nothing but eminently sensible. And much critical response to Rawls operates along lines and assumptions not dissimilar to the above. Rawls relies heavily on Sen, as the numerous references in the index attest. But as most readers of Rawls realize, all of the above is directly opposed to Rawls's assumptions and path of system-building. Rawls for the original position postulates *disinterest* in others. He excludes benevolence. In the "decision-space" behind the veil in the original position, no individuals at all are present.[63] Sen indicates that the "insular" economic man is insufficiently dimensioned, and too weakly connected to the values of his or her society to qualify as a centerpiece for social choice. Rawls on the other hand embraces this level of abstraction. He finds, exactly where Sen feels he cannot, a fruitful way to frame and manipulate his problem. He moves — not away from economics as Sen advises — but towards it.

> Ideally the rules should be set up in ways which further socially desirable ends. The conduct of individuals guided by their rational plans should be coordinated as far as possible to achieve results which although not intended or perhaps even foreseen by them are nevertheless the best ones from the standpoint of social justice. (*Theory* 57; *TJR* 49)

[63] Rawls vacillates in his descriptions, stating and then correcting. This variation has empowered critics to choose one description over another, since the text apparently permits them all. It is the argument here that this is a misstep. One example of many: in the "Classical Utilitarianism" section, Rawls states, "Although justice as fairness begins by taking the persons in the original position as individuals, or more accurately as continuing strands…", here correcting the idea of individuals with something less than fully human, a continuing strand. Rawls isn't ambiguous about which description is more accurate. (*Theory* 192; *TJR* 166)

Excluding the achieving of social justice, this is almost exactly the classic definition of markets.

Rawls's system separates motivation from outcome. Motivation in the original position, as we will see, is not moral, and yet its outcome is. This is major shift away from the utilitarianism that preceded him — where moral *intention* is paramount — to accept that motivation without overt moral content can produce what is essentially a moral result.[64] This is a lesson from economics: as with self-interest in the market generating beneficial social outcomes, so Rawls's system, lacking direct interest in others and absent benevolence, can generate an outcome in which those dimensions can flourish. This gives some indication of what a paradigm-shift the assumptions behind Rawls's theory are.

One might initially wonder whether the informational problem Sen perceives is amenable to a sort of compromise, or shorthand, assessment. Perhaps "full information" is an ideal benchmark, but some approximation on the margin, given the constraints and restrictions, is sufficient. Elster reminds us that such a compromise exists. "Satisficing" was Herbert Simon's compound word (satisfy and suffice) to denote a decision that is not fully optimal, because of difficulty processing information, or the misleading nature of some information. The agent has, not a sense of needing complete information, but an aspirational threshold . This is a line, or demarcation, above which is satisfactory, and below which is not. But Simon's ideal state would still involve full information. It is just

[64] "For the fact that in the original position the parties are characterized as mutually disinterested does not entail that persons in ordinary life or in a well-ordered society, who hold the principles that would be agreed to are similarly disinterested in one another." (*Theory* 147-8; *TJR* 128)

that real world computational or perceptual problems generally put such a possibility out of reach.[65] This compromise notion, however, is not at all relevant to the issue at hand in Rawls, nor to similar concerns in the inconsistent plans literature. In those cases, full information is to be avoided, not approximated. The argument there is that certain sorts of information *negatively* affect decisions. Shorthand solutions like satisficing don't address the problems being considered.

Before moving to particulars, we should briefly look at what Rawls means by "fairness" in his scheme. Fairness, as he points out, does not pertain to outcomes, but to the process by which the principles emerge. Individuals without certain strands of knowledge are in a "fair" position vis-a-vis one another. This is in line with the overall shape of the theory as a pure procedure. A pure procedure is one designed in such a way that the output will be consistent with the procedure. But, critically, this contrasts with a perfect procedure, where exactly what the result should be is known in advance. A perfect procedure, then, is subject to verification; we can see if the output matches our idea of what it should be. In Rawls's example of cutting a cake, we want the slices to be equal-sized. But each individual is assumed to want the largest slice possible. The procedure is to let one person cut, and the other choose. The procedure generates what we intend: equal-sized slices. A pure procedure has no such verification. Rawls's example is gambling — we agree on the how, but not on the final result. The *result* of a pure procedure would be self-justifying; it is the design of the procedure itself, not its output,

[65] Rawls references Simon as applying to Sidgwick's view of agents having full information and making no predictive errors. Simon is introduced as a way in which reality might intrude on a pure deliberative procedure. (*Theory* 416ff; *TJR* 365ff.)

which should be open to scrutiny. Principles of justice in Rawls's theory, he tells us, are arrived at through a pure procedure.

The initial conditions (in the original position) are characterized by fairness. And Rawls lets this idea of what is fair be the guide as to what information will be deemed admissible behind the veil. Fairness is characterized by what information, what self-knowledge, is to be allowed to operate in the original position. The restrictions block certain arguments for the principles of justice, which might in turn affect the principles themselves (*Theory* 18; *TJR* 16).[66] For example, it would not be fair, Rawls argues, for wealthy individuals to be aware of their wealth, lest they naturally favor lower sorts of taxes. Similarly, the poor, knowing their position, might argue the opposite. Either side triumphing could be considered an unfair outcome. Rawls seeks to prune away these types of unfairness, which are sources of disagreement. His argument is, if initial conditions are fair in this sense, rational deliberation will flow into a single channel of generality. But how much informational filtering is required to accomplish this? One key is that we are dealing with a pure procedure. The situation is created —only information of certain sorts is permitted — and once that is set, the pure procedure operates like a machine.[67] But there appears some confusion in Rawls's argument here. *His* system does operate to generate a single outcome — this is his claim. Yet his example of gambling

[66] It is to be "a status quo in which any agreements reached are fair." (*Theory* 120; *TJR* 104)
[67] The machine metaphor is not uncommon in *Theory*. It appears even in his description of politics, "We may think of the political process as a machine which makes social decisions when the views of representatives and their constituents are fed into it." (*Theory* 196; *TJR* 171-2)

does not do this. The difference seems to be the element of chance; Rawls seeks to exclude chance and its attendant risk-assessments from his system.[68]

We have so far seen that both Rawls's system and the inconsistent plans literature fashion their baselines of rationality by characterizing their initial decision-space as one where information is filtered. These informational restrictions are assumed to allow deciders (or "deciders" in Rawls's case) to reach better decisions. And as background to this arrangement, we have looked at Rawls's use of markets as models. Rawls tells us he seeks a pure procedure. The market, he argues, is a perfect procedure for efficiency (*Theory* 359; *TJR* 316). Agents in market activity (in ideal theory) are rational, but certainly not completely dimensional as human beings. This allows their rationality to be fully predictable (in the pure theory); it has an automatic nature. I argue that Rawls aims to replicate this automatic dimension for his social theory. "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". The idea is to "avoid introducing into it any controversial moral elements" (*Theory* 14; *TJR* 12). This appears straightforward. Avoiding the moral discriminations and gradations found in the utilitarianism of Mill or Edgeworth (or Sidgwick), the honed rationality operational in Rawls (honed through informational restrictions) is *designed* to produce or output the principles of Rawls's system.[69] Informational filtering, then, shapes the rational impulse,

[68] Gambling is not really a helpful example.

[69] Hardin (1986) argues persuasively that Mill's utilitarianism is mischaracterized by calling it rule-utilitarianism. Hardin feels it is more appropriately seen as institutional-utilitarianism. The proximity to Rawls's thinking is marked: "The primary subject of the principles of social justice

which *inevitably* generates the stipulated hierarchy of principles. "The theory of justice is a part, perhaps the most significant part, of the theory of rational choice" (*Theory* 16; *TJR* 15).

II. Background

We begin with a question: why haven't this inconsistent plans literature and Rawls been brought into conjunction before? One possibility is their differing emphases. And these emphases lie in two areas that Elster claims differentiate natural selection and human behavior: the capacities for *global maximization* and for *strategic behavior* (Elster 1984, 2). For Elster it is intentionality that sets these two ideas apart from activity in nature. In our smaller context, global maximization connotes a plan which we never repudiate, one which we believe best through all considered time periods. Behavior which deviates is therefore considered simply a plan violation.[70] Whether there are decision-environments in which such globally maximal plans are visible in greater relief, are plainer to us, is one of the central questions to be addressed. Strategic behavior, then, assumes the globally maximal nature of a plan. Given this, violations of such a plan will then be regretted, and should be strategized against. The word used in this context — precommitment — establishes the separation in time. "To bind oneself is to carry out a

---

is the basic structure of society, the arrangement of major social institutions into one scheme of cooperation." (*Theory* 54; *TJR* 47)

[70] Thus, more precisely for our purposes, "For intentional adaptation… we do have a general mechanism for attaining global maxima, and what needs a separate explanation is rather the failure to achieve this." (Elster 1984, 3)

certain decision at time t_1 in order to increase the probability that one will carry out another decision at time t_2" (Elster 1979, 39). These ideas span both *Theory* and the inconsistent plan literature. But they are not equally central. The inconsistent plans literature has a strong focus on the strategic dimension, on "keeping to the plan". This involves forms of self-constraint, and confronts the problem of inclinations at war with one another. And it outlines strategies, such as

> … to stop smoking it is standard practice to set up some causal machinery that will add force to your inner resolution: to tell your friends about your intention so as to invite their sarcastic comments if you are backsliding; to go for a walk in the mountains so as to make cigarettes physically unavailable, … to undergo hypnosis in order to induce aversion to tobacco; to make yourself believe that more cigarettes mean certain death within five years (etc.). (Elster 1979, 37)

— as solutions to a problem. Global maximization — finding a best consistent plan, and deciding on which decision-environment most conduces to making it — is described. But it is not the problem's locus.

These emphases are reversed for Rawls. In *Theory* the critical concern is global maximization; how are principles to be derived. In some ways, Rawls sets this up as a traditional maximization problem: how, subject to certain constraints, can we identify a maximum? In that sense, all of the "action" is in the constraints. The decision environment (the original position) contains those constraints. And those particular constraints are informational. While also important for the inconsistent plans literature, the process of plan formulation is pivotal for Rawls. What is the decision-environment

like, what information is allowed and what is not, are the key questions in the dynamic that will produce his principles. The stability of the principles once individuals emerge from behind the veil -- the strategic elements – are of course also critical. But the types of strategies differ somewhat from those we find with inconsistent plans. But Rawls still asks: will those individuals honor the principles still? How strong will the temptation to violate them be, and why? What can be done to increase the likelihood of the principles being honored?

The inconsistent plans literature overflows with examples and particularities. The situations are posed as problems; the analysis attempts to isolate solutions. If one has trouble saving money, how are Christmas Clubs (now out of fashion) or having extra tax taken from one's wages solutions? In that sense they are practical in nature, in a way that Rawls's theory is frequently viewed as essentially practical (Audard 2007, 22; Freeman 2007, 187).[71] We ask: what is the proper vantage point from which the best decisions are made? What problems with everyday decisions do such changes of perspective address? In what sense can we privilege either perspective (the distant or the immediate), and what is entailed in evaluating any decision as superior? Studies of self-command, self-deception, addiction, time-discounting, resolution-keeping, and a host of others all engage with these problems. I am utilizing them as a single sort of study, under the umbrella term of plan-inconsistency. But despite the informational and strategic overlap of inconsistent plans and *A Theory of Justice,* still this connection may need to be solidified. Thus, why, if these parallels are informative, do we not find more

[71] "Rawls conceives of the primary purpose of political philosophy in a democratic society as practical (as opposed to epistemological or metaphysical)…." (Freeman 2007, 187)

terminological overlap between them? Why don't Rawls's prodigious footnotes reflect a familiarity with this literature? There are various answers to these questions. What one might call the ur-paper for these questions, (Strotz, 1955-6), although early enough for Rawls to have seen, did not evoke a significant response at the time. The groundswell of interest — led by Elster, Thaler, Shefrin, Ainslie, Schelling, and others — came largely after *Theory*.[72] If *Theory* had emerged ten years later, it is interesting to speculate on what Rawls's reaction to these papers would have been. On his own, he covered some topics to which they refer: time-discounting and competing selves. The papers are also involved with questions of rationality, which are at the very heart of *Theory.* We will see how these various concerns illuminate one another.

III. Informational Restrictions

The argument here is that Rawls's severe restrictions on information in the original position are, with rationality, the critical elements in Rawls's ability to isolate his principles.[73] The inconsistent plans literature can help us see Rawls's construction as part of a more general solution-type. We saw that Elster designated two times — t_1 and t_2 — to differentiate the decision moment, from the action (or temptation) moment. This has two presumed advantages. First it views the action state from a distance in time, lessening the immediate features which might tempt deviation from the plan later. Secondly, there

[72] There is an open question why Rawls's indebtedness to economics was not continued in subsequent writings. But whatever the cause, it does seem to have clouded his earlier indebtedness for many commentators.

[73] "To say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion." (*Theory* 138; *TJR* 119-120)

is most likely physical distance, which serves the same purpose. It is the first of these that psychologist George Ainslie has in mind when he says "a larger, later reward is preferred when the choice is seen from a distance, but the smaller, earlier reward is preferred as it becomes imminent" (1985, 141-2). The two advantages are identical in their effect; each takes the immediate circumstances of a later moment, and reduces the effect such circumstances have on the decision at time t_1. The problem that calls for this kind of thinking is usefully described by Schelling,

> Specifically, if I could decide now not to eat dessert at dinner, not to smoke a cigarette with my coffee, not to have a second glass of wine, and not to watch the late movie after I get home, I would make those decisions because *now* I want *not* to do those things *then.* And I know that when the time arrives I shall want to do those things and will do them. I now prefer to frustrate my later preferences. (1983, 85)

Schelling's overall plan — the "nots" — can be seen as a small-world global maximum. It is assumed there will not be a moment where the plan itself will be abandoned. He will not change his mind as he eats his dessert, and claim this option is better, and that the plan was foolish. The plan, in this scenario, remains intact. He will regret not following it.

We are assuming, then, for the moment that these long-range plans, taken previous to actual activity, are *best.* And we are assuming as well that the situations which pose a future threat to these plans are anticipated. We then run through potential scenarios where the plans are violated and where they are not, and prefer that they are

not. Strotz (1955-6) presents the problem as when these made-in-advance plans actually prove out, and "original expectations of future desires and means of consumption are verified" (165). Surprises are outside of the problem specification. We see this framing of the problem largely matches that in *Theory.* Temptations gain power from proximity; greater distance from them favors choosing the other alternative, one based on longer range concerns. In terms of examples, Ulysses does not wait until he hears the Sirens' call to decide whether to have himself bound to the mast; the alcoholic does not go to a bar to formulate his resolution to cease drinking. In these examples and others, immediate circumstances, *although increasing the information set of the decider*, do not contribute to the best decision-making process. In the decision phase, all authors with this view describe decisions made *not* within *proximity* of these less satisfactory choices. The goal for decision-making, then, is some degree of sequestration.

The individual[74] in *Theory* is sequestered as well, but to a much greater extent. A famous passage details the categories,

> Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. (*Theory* 12; *TJR* 11)

[74] I am calling what makes decisions behind the veil an individual here, but as we will see, no individuals are in fact involved.

The paucity of information in the original position goes to the heart of Rawls's theory. He tells us these restrictions are of "fundamental importance". "Without them we would not be able to work out any definite theory of justice at all" (*Theory* 140; *TJR* 121). At first glance, we can see the requirement for these restrictions in contrast to the inconsistent plans articles. In those articles, only the circumstance relevant to the decision is to be distanced. A vow to avoid bakeries might well be taken in a bar; a vow to avoid bars might be taken in a bakery. The separation in time is present, but need not necessarily be substantial. Larger issues may require deeper sequestration. Deciding on divorce might best be done during time apart, in a state of calm. Comparatively, what Rawls's agent is deciding is enormous: nothing less than the principles to structure society's institutions. We can extrapolate out how extensive the isolation, and restriction on information, might have to be.

The greater complexity of the problem in *Theory,* however, isn't only a question of *more*. To say that more information is restricted in the original position simply because the problem at hand is more profound doesn't completely help us. Individuals in the inconsistent plans examples retain their individuality. Their problems are personal, and their intimate acquaintance with them and their particular psychology linked to those problems are intact. This is emphatically not the case in *Theory.* And, of course, with inconsistent plans, any number of possible strategies or plans might result. There is no sense that a certain decision-environment will determine a particular plan. But this is precisely what Rawls does claim. To achieve even a semblance of this result a number of different factors must come together. This white-hot concentration has been greeted

skeptically by many readers, but it has also been misunderstood. Sandel (1982) can stand in for many,

> Discussion, like bargaining, presupposes some differences in the perceptions or interest or knowledge or concerns of the discussants, but in the original position there are no such differences. We must therefore assume that the 'deliberations' of the parties proceed in silence and issue in a single conception which is unanimously agreed to.
>
> But this makes the account of the agreement in the original position more puzzling still. For if there is no basis for bargaining or discussion, it is doubtful that there can be any basis for agreement, even a unanimous agreement, either. (129)

Here, the assumption of actual individuals behind the veil leads him astray. Agreement, so-called, Rawls informs us, results from the informational contouring. The principles don't arise from exchange across individuals. The bargaining aspect is particularly to be avoided, as it requires information and involves power-relationships.[75] The attempt to more fully animate the original position, past what Rawls indicates, leads to unnecessary confusion.

Rawls at one point says it is as if one person only is behind the veil. What this implies is that all individual differences, relevant to this particular choice, have been eradicated. It is the informational filtering that has accomplished this result. Rawls also

[75] Rawls adopts a view on bargaining that parallels one of his major influences, Frank Knight. (*Theory* 139; *TJR* 120-1) See Coker (2021) for a fleshing out of this relationship.

says at numerous points that the "decision" is purely deductive.[76] There is therefore only one possible choice. We may not like this sort of procedure, but it does accomplish its goal. Like actors in perfect competition faced with a price change, all other factors *ceteris paribus*, the "decision" is actually non-existent. There is only the assumption. Rawls claims that his individuals do have "a certain psychology", since very general descriptions of their beliefs are in play. But to discuss individuals in this context is only a manner of speaking. It is hard, descriptively, to avoid. This is so particularly as Rawls encourages the abstract possibility of assuming the conditions he specifies, to check the reasonableness of the generated principles. But nothing, he notes, hinges on whether this is actually feasible. This is so because the mechanism is justified on other grounds. Rawls, as he so frequently does, references economic parallels at this crucial juncture,

> These assumptions 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. (*Theory* 121; *TJR* 105), emphasis added)

[76] "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 their acknowledgment is the only choice consistent with the full description of the original position. The argument aims eventually to be strictly deductive." (*Theory* 121; *TJR* 105)

Clearly, in a world of moral geometry, nothing is decided by discussion. All rational application winds up with the same answer. It is the system which determines the answer, not particular individuals.

Rawls's response to utilitarianism provides an additional view of information. Sidgwick's version of the theory — Rawls's primary reference[77] — posits an impartial observer. This observer must display two characteristics (or powers): it must have full information, and be fully sympathetic (benevolent). Only in this way can society be assured that utility valuations will be weighed accurately and to best advantage. Rawls's system eliminates both of these traits. As with his disagreement with Sen, Rawls argues that utilitarianism's goals can best be reached through a very different set of assumptions. Again he puts rational self-interest — in a carefully configured decision-environment — against a seemingly more humane assumption of information plus benevolence. Rawls insists that utilitarianism assigns concern for others to its participants. Benevolence must be considered a duty.[78] Rawls says his system avoids duties — in that way it is closer to a rights-based theory than utilitarianism. He further claims that his veil of ignorance plus mutual disinterestedness approximate benevolence.[79] And they do so without the

---

[77] "I shall take Henry Sidgwick's *The Methods of Ethics* as summarizing the development of utilitarian moral theory." (*Theory* 22; *TJR* 20) And Sidgwick acknowledges the problem of inconsistent plans: "As a species intermediate between the two, we may place resolutions to act in a certain way at some future time: we continually make such resolutions and sometimes when the time comes for carrying them out, we do in fact act otherwise under the influence of passion or mere habit, without consciously cancelling our previous resolve. This inconsistency of will our practical reason condemns as irrational…." (Sidgwick 1981, 37)

[78]" Here I wish only to point out that if the duty of aiming at the general happiness is the taken to include all other duties, as subordinate applications of it, we seem to be again led to the notion of Happiness as an ultimate end categorically prescribed…." (Sidgwick, 8)

[79] Mill also discusses duty in his "Utilitarianism". He associates it particularly with justice, and explains the coercive dimension that must accompany it. "It is a part of the notion of Duty in

drawbacks that utilitarianism entails. Justice, we may recall, for Rawls is oppositional. Benevolence therefore cannot be foundational.[80] And benevolence here represents moral concern generally — again barred from being foundational. Morality is not a decisive force behind the veil (Audard 2007, 127; Freeman 2007, 151). And benevolence, besides, introduces a variety of attitudes and assessments that will resist reconciliation. Whose idea of benevolence will dominate? Full information, on the other hand, obstructs unanimity in that it entails confusion. Utilitarianism introduces a mythical spectator for a reason: full information is beyond the capacity of actual individuals to assimilate. Rawls postpones his concern for aspects such as benevolence until after his principles emerge. Justice for Rawls is primary; in a just world the finer dimensions of human impulse can flourish. But benevolence plus knowledge as primary constituents are

> … so complex that no definite theory at all can be worked out, Not only are the complications caused by so much information insurmountable, but the motivational assumption requires clarification. For example, what is the relative strength of benevolent desires? In brief, the combination of mutual disinterestedness plus the veil of ignorance has the merits of simplicity and clarity while at the same insuring the effects of what are at first sight morally more attractive assumptions. (*Theory* 148; *TJR* 129)

---

every one of its forms, that a person may rightfully be compelled to fulfil it. Duty is a thing which may be *exacted* from a person, as one exacts a debt." (246) Contract theorists, such as Rawls and Buchanan, aim particularly to avoid such coercion.

[80] It probably did not escape Rawls's notice that benevolence is also excluded explicitly in *The Wealth of Nations.* In perhaps its most famous passage, "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest" (Smith 26-7). Economics is not a benevolence-based discipline.

Rationality on the other hand, anchored in the original position, makes a virtue of lack of knowledge. Reduced to general concerns, looking to "the advancement of human interests broadly defined", the self-interest (or maximizing) impulse moves directly towards its social target.

So "individuals" behind the veil are not morally motivated, they are simply rational maximizers. The economic model hovering just behind this formulation is not obscure. Agents in a market setting, through rational self-interest, produce unintended beneficial outcomes. In both systems, motivation is distinct from output, with rational self-interest being *situated* in environments (the original position/the market) that guides actions towards beneficial results. Introducing moral factors in Rawls would work against unanimity, and create potential grounds for disagreement. Without such factors, a streamlined rationality is predictable. And this predictability is so precise that it operates irrespective of which agent's mind it is in — that is, we can abstract from individuals altogether.

And actors in the marketplace also operate in a diminished informational setting. It is not necessary, perhaps not even helpful, to know anything personal about the person one is dealing with. How many children they have, where they like to eat, what color their car is are all irrelevant for market interactions in economic theory. From the standpoint of theory, such transactions are essentially "faceless". At this level of abstraction, the market is merely a mechanism. I suggest that the *automatic-ness* of such markets is a deep model in Rawls's system. And participants in such abstract markets not

needing enormous levels of information is key to their functionality.[81] In theory, market decisions don't pivot on such information. To include such factors would muddy the simple drive toward general outcomes. And Rawls's use of rationality, and exclusion of moral considerations as motivation, follows the market pattern. Narrow rational motivation can lead to beneficial social outcomes. His theory will try to duplicate this trajectory. He thereby steps away from the moral intricacies and indeterminacies that Sidgwick's (and Mill's) utilitarianism must confront, and even from Sen's attempt to add (what he considers) relevant dimensions to (what he considers) narrow economic calculation. It is this core drive that more than anything establishes Rawls's originality. The mechanical nature of his original position, despite the often human face it assumes for expository purposes, is central. And its ability to generate his principles relies on the carefully curated informational environment he specifies.

IV. Rationality

We can look at rationality using the idea of global and local maxima. In the inconsistent plans literature, it isn't as if what constitutes temptation has nothing to recommend it. If this were so, there wouldn't be a problem to start with. No one has to self-constrain their future self from the temptation of driving a nail through their foot. Such temptations tend to take care of themselves. Temptations to be guarded against have some utility benefit. They are simply, when considered from a certain perspective, not the best option for long-run gain. They are, however, within their limited time frame, an

[81] I am going to explain this on the personal level, but certainly Hayek's informational arguments are germane here as well (Hayek 1945).

option which might be selected; they are local maxima. The problem in that literature, then, has two dimensions. One is to hold to these long-run plans through whatever strategies are available. The other is to find or create the decision-environment from which global maxima can best be distinguished from local ones. Rawls's dilemma looks remarkably similar. Agents in the real world, aware of the advantages to be gained in short-run or short-term activities, potentially have difficulty with behavioral ideals. But Rawls begins at a slightly different point: he wonders whether individuals, in the midst of these temptations, are properly situated to even envision where precisely the global maxima are — or, better, what principles could best clear a path leading in their direction. The oppositional nature of everyday striving, even considering the substantial cooperative dimensions which are present as well, make attaining a long-run perspective a considerable challenge. There is a strong sense in *Theory* that the difficulty level may not just be high, but unnecessarily high. Could we craft an environment which would help us see ourselves in relation to global maxima more clearly?

Envisioned in this way, Rawls's strategic solution appears straightforward. The everyday world is a disturbing mixture of global and local maxima; the presence of one obscures our evaluation of the other. Self-interest thus situated displays a blend of advantageous and disadvantageous factors. Systems like utilitarianism array a number of other-interested inclinations to influence this self-interest towards its more constructive side. But the benevolence Sidgwick supposes to be operational clearly cannot capture self-interest entirely. What it can't co-opt, it hopes to suppress. The inconsistent plans literature more overtly fails to control certain short-term impulses — it focuses on the

ways in which sterner measures must be taken to ensure results. But Rawls points to a self-interested behavior that leads to positive outcomes — the market. It does so, however, only because it is set up in a certain manner. On a theoretical level the operation is shown in even greater relief. There the *model* of the market characterizes individuals in only a few dimensions. This lack of dimensionality lets the self-interested rationality of the actor become highly predictable. The model becomes a world of assumption, a world of correct anticipation on the part of the theorist.[82] It takes serious liberties with reality as we know it. It is Rawls's great insight to see the potential this all has for social theory, for discovering principles that set institutional limits.

The inconsistent plans literature helps us see Rawls's use of rationality somewhat more clearly. In particular, Strotz tells us (as we saw earlier) that rationality characterizes a plan that we hold to through all relevant time periods. Under that assumption, violating the plan is *irrational.* This is Rawls's contention as well — agents are thought to formulate plans they will not renounce. And this also comes under the heading of rationality. This brings us to two problems Rawls must answer. The general problem is: what of not following the plan when emerging from the veil of ignorance? And this in turn has two possibilities: the individual can either decide the plan was in error, or she can decide to violate the plan for personal advantage.

One way the plan can be violated is that the agent simply decides at time t_2 that the near-field temptation is now the correct choice. But this is not a violation of a plan still held, it is rather a reevaluation given new circumstances. The plan is believed

[82] Gilboa, Postlewaite, and Schmeidler argue that rationality for an agent requires she be aware of the "right" model, "that is, the one used by the external observer." (2012, 27)

mistaken, at least within that particular time period. Theorists like Thaler and Shefrin (1981) model time decision-disparities as a divided self, but this is merely an expository device, without a psychological claim. Schelling (1980) on the other hand views this division as potentially real. Flip-flopping behavior might be one indicator: if the immediate self "eats a high-calorie lunch knowing that he will regret it, does regret it, cannot understand how he lost control, resolves to compensate with a low-calorie dinner, eats a high-calorie dinner knowing that he will regret it, and does regret it", then we have two perspectives, perhaps, but one (the one suffering regret) is probably privileged. Other examples are less clear: which self is privileged if, after a drink too many, [one empties] one's wallet into the Salvation Army bucket?[83] Strotz avoids this problem by postulating a plan that attains the individual's approval through all periods, even if the behavior deviates from it. Schelling's question, therefore — about whether a plan should be privileged — is also an attack on the assumption of rationality.[84] If the plan is not "best" (or even good), holding to it is surely not a sound definition of rationality. Schelling further folds the dimensional space by suggesting that, if Arrow[85] has proved that the individual is not a good model for collective decision-making, perhaps the collective is a more appropriate model for individual decision-making. We are, in that sense, a number of successive individuals. And successive individuals are almost certainly prone to behaving inconsistently— an outcome Rawls is determined to sidestep. For Schelling, the

[83] Schelling 80, 59ff. in *Choice and Consequence.*
[84] "So we should not expect a person's choices on the matters that give rise to alternating values to display the qualities typically imputed to rational decision, like transitivity, irrelevance of 'irrelevant' alternatives, and short-run stability over time." (Schelling 83, 94 in *Choice and Consequence.*)
[85] The implied reference is to Arrow's *Social Choice and Individual Values* (see references).

possibility of multiple selves introduces the idea of multiple value systems — another concept Rawls would avoid.

This is the first prong of the problem we want to look at: the individual either sporadically or permanently finds the original plan to be in error. The sporadic option is easily addressed — Rawls insists that his individuals should not favor one time period over another. Time preference is to be rejected.[86] Here he follows not only Sidgwick, but also Frank Ramsey (1928). [Ramsey is also discussed by Strotz — the closest connection between Rawls and the literature under discussion.] Rawls's argument here is one of symmetry; there is no more reason for the present person to discount the future, than for that future person to discount the present. Rawls intends his individuals to imagine themselves at all of these positions. And when the question is generational, Rawls simply says that privileging one generation at the expense of another is to let some flourish at the expense of others — his original argument against utilitarianism.

The second variant is that one comes to find, permanently, that the decision behind the veil was an error. This is, indirectly, a critique of Rawls's system in general; why should we as readers believe that the two principles are in fact what would emerge in the original position? Perhaps, from our location outside the veil, we already feel this might not be the case. Rawls argues that it is possible that unforeseen circumstances could make us regret our choice. Factors we could not have anticipated could intrude. Or, it could *be* the best plan, but all possible plans turn out to be abhorrent. In such a case,

[86] "We are to see our life as one whole, the activities of one rational subject spread out in time. Mere temporal position, or distance from the present, is not a reason for favoring one moment over another." (*Theory* 420; *TJR* 369)

Rawls says we might "wish that we had never been born" (*Theory* 422; *TJR* 370). But we can have no regret *about having chosen the plan we did.* The choice itself is beyond regret. And on this elusive point we see his system cohere: the rationality operational in the original position can only select one output. It is not the choice which might swerve in different directions; that is automatic. It is the circumstances of the choice that are open to critique. The original position is "a natural guide to intuition", nothing more. No persons inhabit it. It is a perspective that "one can at any time adopt".

> It must make no difference when one takes up this viewpoint, or who does so; the restrictions must be such that the same principles are always chosen. The veil of ignorance is a key condition in meeting this requirement. (*Theory* 139; *TJR* 120)

Individuals who must always make a certain choice, given certain circumstances, are abstractions. If the choice is mechanical, who or what makes it is irrelevant.

We can be brief with the second possibility: that actors deliberately violate the plan for self-interested reasons. This is the problem framed in Hobbes, and Hume.[87] Rawls analyzes the possibilities in "The Sense of Justice" (1963), where the defector option has negative externalities, serving to erode trust. Given this possibility, Rawls is concerned like Hobbes to bring constraining factors to bear. Hobbes's presentation of the situation Rawls links to the prisoner's dilemma problem (explicitly in *Theory* 269; *TJR* 238).[88] In that problem, the defection option garners the highest payoff, no matter the choice of the other player. Hobbes's answer to this is to make sure the defection option is

---

[87] *Leviathan* and *A Treatise of Human Nature,* 488ff.

[88] And the references around Rawls's discussion of Hobbes are not philosophical, but economic and game-theoretic: W.J. Baumol's *Welfare Economics and the Theory of the State* and R.D. Luce and Howard Raffia's *Games and Decisions*.

not selected. Rawls of course does not follow Hobbes and his concept of a central authority wielding a coercive threat. Instead, Rawls discusses something which is unusual considering the view taken in this paper; he discusses guilt. Guilt as a constraining force is developed both in this paper, and in *Theory*, at some length. But what is of interest here is *why* the individual is vulnerable to this constraint. The reason is that, to knowingly violate an agreement such as the one Rawls stipulates, is to violate as well one's own sense of justice. Here we find a supposition that Rawls takes exceedingly seriously.

> Thus a person who lacks a sense of justice is also without certain natural attitudes and certain moral feelings of a particularly elementary kind. Put another way, one who lacks a sense of justice lacks certain fundamental attitudes and capacities included under the notion of humanity. (Rawls 1963, 111 in *Collected Papers*)

It is hard to imagine a more severe rebuke. Rawls tries to build his system assuming as little as possible, but he assumes this. Being human implies having an inborn sense of justice.[89] It is as foundational as Rawls will allow himself to be.

But Rawls has already told us that moral arguments are not influencing decisions in the original position — agents are not characterized as benevolent or interested in others. So where does the sense of justice fit behind the veil? Our reason for the enterprise as a whole — to find principles of justice — does require that we be concerned with justice. So its absence would make the entire enterprise meaningless. We note, however, that the sense of justice is not involved in actually selecting the principles.

---

[89] Rawls offers that it is "almost certain that at least the vast majority of mankind have a capacity for a sense of justice and that, for all practical purposes, one may safely assume that all men originally possess it." (1963, 114 in *Collected Papers*)

Moral arguments do not do that — that decision is a maximizing exercise. Deciding on the principles Rawls leaves to his carefully contoured rationality. Rationality is the key operating force in the pure procedure that is Rawls's theory.

We can see how this complicated skein of ideas — informational constraints fashioning rationality, the market as a perfect procedural model, the automatic nature of the "process" that outputs Rawls's principles — all work together, and are coherent. Their coherence gives us perspective to answer some criticisms of the system, criticisms that ignore one or more of these features. One such is the idea that Rawls's references to a social contract are merely distracting; in fact, no one has ever explicitly signed on to this contract (Dworkin, in Larmore 370).[90] Rawls of course admits this obvious point. Rawls defends his use of contract by saying that one *agrees* to the stipulations of the original position. These "we do in fact accept" (*Theory* 21; *TJR* 19). And this is the only voluntaristic step required. If we also recognize the automatic nature of decisions in the original position, as has been argued here, then the principles are accepted when the original position is accepted. And the contract form can be seen as emphasizing the non-foundational nature of the enterprise. The contract is a coming-together of viewpoints, but this is achieved through a particular convergence of rational considerations, driven in a certain direction by the informational environment in which it operates.[91]

---

[90] Criticized of course in Mill (2006, 252-3) and Hume (1987/1777).

[91] It should additionally be mentioned that Rawls feels that contract theory has analytical advantages. By varying initial conditions — on which outcomes depend — it is clearer which sorts of information link to which sort of social theory. See *Theory* 121-2; *TJR* 105.

One can also consider the criticism that Rawls's "individuals" are too thinly rendered to be able to formulate the stated conclusions. Someone as astute as Gauthier says,

> In seeking to treat persons as pure beings freed from the arbitrariness of their individuating characteristics, Rawls succeeds in treating persons only as social instruments. … In his argument morality is divorced from the standpoint of the individual actor.
>
> [As with Rawls's charge that utilitarianism fails to take] "seriously the distinction between persons… Rawls's theory falls victim to the same charge." (Gauthier 254)

Rawls's objection that Gauthier cites is that in utilitarianism, it is possible for the interests of some to be sacrificed to the interests of others. But the wider response to this is, simply, yes, Rawls does do this. But to criticize it is to misunderstand the entire process at work in his theory. Individuals behind the veil are indeed "instruments".[92] They are, after all, part of a pure procedure, cogs in that mechanism. Rawls's original position is a thought experiment only secondarily. Nothing hinges on the success or failure of that experiment. He explicitly says that failure to see the principles when screening oneself from personal information (in the imagination) implies nothing about the certainties involved in the formulation.

[92] Barber's witticism, that disinterested individuals might turn out to be uninterested individuals, is off the mark for the same reason. But it is amusing. (Barber 1974, 295)

V. Conclusion

This notion of a procedure is in many respects the engine at the center of Rawls's *Theory*. "The aim is to use the notion of pure procedural justice as a basis of theory" (*Theory* 136; *TJR3* 118). Rational self-interest is to be placed in such a carefully constructed situation, that letting natural inclinations play out, the result is (in our cake example) *as if* fairness had been the driving criterion. The huge insight for Rawls was that this was relevant to social philosophy. If one could construct such a procedure, many of the roadblock problems of value and evaluation could be sidestepped. And there was an ideal theory ready to hand to use as a model: the theory of the market. As noted earlier, for Rawls, the market is a perfect procedure for efficiency. He clearly views market outcomes as a general good resulting from individualized acts of self-interest, and intends us to see the connection to his system. And, as with the market, his system disassociates motivation from outcome. Rationality, not morality, drives both systems. In the section in *Theory* on "Economic Systems" (Section 42), after detailing ideal market operations, and then cataloguing some imperfections that might muddy those results, he says, "But these matters need not concern us here. These idealized arrangements are mentioned in order to clarify the related notion of pure procedural justice" (*Theory* 272; *TJR* 240).

Informational filtering is the major ingredient in creating the space in which this pure procedure can function. We have seen how such filtering has worked in the related problems of plan inconsistency. Tying two strands together — use of the market as model, and the family resemblance to the plan inconsistency literature — we see that, as

opposed to the original position conceived as a sort of flight-of-fancy, a proposition that strains credulity, it is instead an idea anchored in a deep practicality. But the concept of the original position as fully abstract is difficult to sustain when Rawls then mentions individuals there, "I then assume that the parties are presented with this list…" etc. (*Theory* 122; *TJR* 106). We must rely on his summary passages to lift us out of this duality. One such passage shows Rawls doubting this imaginary entering-in.

> Of course, when we try to conduct ourselves in moral argument as its constraints require, we will presumably find that our deliberation and judgments are influenced by our special inclinations and attitudes. Surely it will prove difficult to correct for our various propensities and aversion in striving to adhere to the conditions of this idealized situation. But none of this affect the contention that in the original position rational persons so characterized would make a certain decision. This proposition belongs the theory of justice. (*Theory* 147; *TJR* 127)

This is quite explicit: should the original position prove inhospitable as a mental space, the theory itself would be unaffected. This is its purely deductive nature, its geometric operation. Information aims the arrow of rationality; at the end of its flight is the destination of Rawls's principles.

We can briefly summarize what the comparison with the inconsistent plans literature has gained for us. On the positive side, it presents the problem to be solved as practical in nature. One wishes to control one's behavior, to (presumably) better one's life. The strategy for doing this is to gain perspective through distance, to mute the immediate circumstances that might pull one in a less satisfactory, short-run direction.

This is exactly Rawls's formula as well. Information is restricted (more severely than in the inconsistent plans instances) to favor the long-run perspective. The closeness of this parallel is reinforced through Rawls's own treatment of similar sorts of concerns: the self consistent through time, the refusal to privilege one "self" over another, and the concern over holding to the principles once the veil is dissolved. The use of information in both the inconsistent plans authors and in Rawls is counter-intuitive; where throughout his system, more is preferred to less (primary goods, etc.), here, less is preferred to more. But in both examples (Rawls and the inconsistent plans literature) the goal is clarity. In that sense, they are all simplifying the real world to bring certain critical features into relief. They are, in short, restricting information to form models.

The negative side is Rawls takes the informational strategy to such an extreme that the parallel is ruptured. Agents in the inconsistent plans literature *are* actually in their decision-space, and they are actually making decisions. I have argued that neither of these is true in Rawls. His unanimity requirement is too demanding; personal latitude cannot be allowed. If it were, his system would have similar drawbacks to the intuitionist and utilitarian views that he hopes to supersede. Rawls is, in a sense, mining for something in human nature that he is sure is there. These are always risky enterprises. An economist as highly sympathetic as Buchanan found it a leap too far. But Buchanan shared what is the theme of this paper: his constitutional framework provided the sort of distanced perspective we have been addressing here. In that sense it has use here as well; showing that Rawls's information strategy is consistent with widespread choice strategies, inside and outside of academic efforts.

REFERENCES

Ainslie, George. (1985). "Beyond microeconomics. Conflict among interests in a multiple self as determinant of value", in Jon Elster, ed., *The Multiple Self.* Cambridge: Cambridge University Press.

Arrow, Kenneth J. (1963). *Social Choice and Individual Values,* 2nd. edition. New York: John Wiley.

Audard, Catherine. (2007). *John Rawls.* Montreal and Kingston: McGill-Queen's University Press.

Barber., Benjamin. (1974). "Justifying Justice: Problems of Psychology, Politics, and Measurement" in Norman Daniels, ed., *Reading Rawls: Critical Studies of* A Theory of Justice. New York: Basic Books.

Baumol, W.J. (1952). *Welfare Economics and the Theory of the State.* London: Longmans, Green.

Buchanan, James M. (1962). "The Relevance of Pareto Optimality", *Journal of Conflict Resolution*, vol. 6, No. 4 (December): 341-354.

Buchanan, James M. and Gordon Tullock. (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy.* Ann Arbor: The University of Michigan Press.

Caldwell, Bruce. (1982). *Beyond Positivism: Economic Methodology in the Twentieth Century.* London: George Allen & Unwin.

Coker, David C. (forthcoming 2021). "Rawls and Knight: Connections and Influence in *A Theory of Justice", Research in the History of Economic Thought and Methodology.*

Elster, Jon. (1977). "Ulysses and the Sirens: A theory of imperfect rationality", *Social Science Information,* Vol. 16, Issue 5: 469-526.

_____. (1984). *Ulysses and the Sirens: Studies in Rationality and Irrationality, revised edition.* Cambridge: Cambridge University Press.

Fleischacker, Samuel. (2004). *A Short History of Distributive Justice.* Cambridge: Harvard University Press.

Forrester, Katrina. (2019). *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy.* Princeton and Oxford: Princeton University Press.

Freeman, Samuel, ed. (2002). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.

_____. (2007). *Rawls*. London and New York: Routledge.

Friedman, Milton. (1953). "The Methodology of Positive Economics", reprinted in *Essays in Positive Economics* (1966). Chicago: The University of Chicago Press (Phoenix Books).

Gauthier, David. (1986). *Morals by Agreement.* Oxford: Oxford University Press.

Gibbard, Allan and Hal Varian. (1978). "Economic Models", *Journal of Philosophy,* 75, 664-677.

Gilboa, Itzhak and Andrew Postlewaite and David Schmeidler. (2012). "Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality", *Synthese*, Vol. 187: 11-31.

Hands, D. Wade. (2001). *Reflection Without Rules: Economic Methodology and Contemporary Science Theory.* Cambridge: Cambridge University Press.

Hardin, Russell. (1986). "The Utilitarian Logic of Liberalism", *Ethics*, Vol. 97, No. 1 (October): 47-74.

Harsanyi, John C. (1977/1990) "Advances in understanding rational behavior" in Paul K. Moser, ed., *Rationality in Action: Contemporary Approaches.* Cambridge: Cambridge University Press.

Hobbes, Thomas. (1651/1968). *Leviathan.* Harmondsworth: Penguin Books.

Knight, Frank H. (1921/1971). *Risk, Uncertainty, and Profit.* Chicago: The University of Chicago Press.

_____. (1923). "The Ethics of Competition", *The Quarterly Journal of Economics*, Vol. 37, Issue 4 (August): 579–624. Reprinted in *The Ethics of Competition.*

_____. (1925). “Economic Psychology and the Value Problem”, *The Quarterly Journal of Economics,* Vol. 39, Issue 3 (May): 372-409. Reprinted in *The Ethics of Competition.*

_____. (1935). *The Ethics of Competition and other essays*. London: George Allen & Unwin Ltd.

_____. (1951). “The Role of Principles in Economics and Politics”, *The American Economic Review,* Vol. 41, No. 1 (March): 1-29. Reprinted in *The Ethics of Competition.*

Kukathas, Chandran and Philip Pettit. (1990). *Rawls: 'A Theory of Justice' and Its Critics*. Palo Alto: Stanford University Press.

Levy, David M. and Sandra J. Peart. (2017). *Escape from Democracy: The Role of Experts and the Public in Economic Policy.* New York, N.Y.: Cambridge University Press.

Luce, R.D. and Howard Raiffa. (1957). *Games and Decisions.* New York: John Wiley and Sons.

Lyons, David. (1974). "Nature and Soundness of the Contract and Coherence Arguments", in *Reading Rawls*. ed. Norman Daniels. New York: Basic Books.

Mandle, Jon. ed. (2013). *A Companion to Rawls*. West Sussex: Wiley Blackwell.

_____. (2009). *Rawls's ‘*A Theory of Justice’*: An Introduction*. Cambridge: Cambridge University Press.

Mill, John Stuart. (2006). *Collected Works of John Stuart Mill, Vol. IV: Essays on Economics and Society 1824-1845.* Indianapolis: Liberty Fund Press.

_____. (2006). “Utilitarianism” in *Collected Works of John Stuart Mill, Vol. 10: Essays on Ethics, Religion and Society.* Indianapolis: Liberty Fund Press.

Morgan, Mary S. (2012). *The World in the Model: How Economists Work and Think.* Cambridge: Cambridge University Press.

Mueller, Dennis C. (1989). *Public Choice II*. Cambridge: Cambridge University Press.

Nagel, Thomas (1974). “Rawls on Justice” in *Reading Rawls: Critical Studies of* A Theory of Justice. New York: Basic Books, Inc.

Nussbaum, Martha C. (2015). "Introduction" in *Rawls's* Political Liberalism, Thom Brooks and Martha C. Nussbaum, eds. New York: Columbia University Press.

O'Neill, Onora. (2015). "Changing Constructions" in *Rawls's* Political Liberalism. Thom Brooks and Martha C. Nussbaum, eds. New York: Columbia University Press.

Patinkin, Don. (1973). "Frank Knight as Teacher", *American Economic Review 63* (December): 787-810.

Peart, Sandra J. and David M. Levy, eds. (2008). *The Street Porter and the Philosopher: Conversations on Analytical Egalitarianism.* Ann Arbor: The University of Michigan Press.

Pogge, Thomas. (2007). *John Rawls: His Life and Theory of Justice.* Oxford: Oxford University Press.

Purcell, Edward A. (1973). *The Crisis of Democratic Theory.* Lexington: The University of Kentucky Press.

Ramsey, F.P. (1928). "A Mathematical Theory of Saving", *The Economic Journal,* Vol. 38, No. 152 (December): 543-559.

Rawls, John. (1958). "Justice as Fairness", *The Philosophical Review,* Vol. 67, No. 2 (April): 164-194; reprinted in *Collected Papers.*

_____. (1963). "The Sense of Justice", *Philosophical Review,* Vol. 72, No. 3 (July): 281-305; reprinted in *Collected Papers.*

_____. (1971). *A Theory of Justice.* Cambridge: The Belknap Press of Harvard University Press.

_____. (1999). *A Theory of Justice, Revised Edition.* The Belknap Press of Harvard University Press.

_____. (1999). *Collected Papers*, ed. Samuel Freeman. Cambridge: Harvard University Press.

_____. (2001). *Justice as Fairness: A Restatement.* Erin Kelly, ed. Cambridge: The Belknap Press of Harvard University Press.

_____. (2005). *Political Liberalism: expanded edition.* New York: Columbia University Press.

_____. (2007). “Of the Original Contract” in *Lectures on the History of Political Philosophy.* Cambridge: The Belknap Press of Harvard University Press.

Reidy, David A. (2014). “From Philosophical Theology to Democratic Theory” in *A Companion to Rawls,* Jon Mandle and David A. Reidy, eds. Malden, MA.: Wiley Blackwell.

Sandel, Michael J. (1998). *Liberalism and the Limits of Justice, second edition.* Cambridge: Cambridge University Press.

Scanlon, T.M. (1974). “Rawls’ *Theory of Justice*”, in *Reading Rawls*, ed. Norman Daniels. New York: Basic Books.

_____. (1982). “Contractualism and Utilitarianism”, in *Utilitarianism and Beyond,* Amartya Sen and Bernard Williams, eds. Cambridge: Cambridge University Press.

_____. (1998). *What We Owe to Each Other*. Cambridge: The Belknap Press of University of Harvard Press.

Schelling, Thomas C. (1980). “The Intimate Contest for Self-Command”, *The Public Interest* 60 (Summer): 94-118. Reprinted in *Choice and Consequence: Perspectives of an errant economist.* Cambridge: Harvard University Press.

_____. (1983). “Ethics, Law, and the Exercise of Self-Command”, in Sterling M. McMurrin, ed. *The Tanner Lectures on Human Values IV*. Salt Lake City: University of Utah Press. Reprinted in *Choice and Consequence: Perspectives of an errant economist.* Cambridge: Harvard University Press.

_____. (1985). “The Mind as a Consuming Organ” in Jon Elster, ed. *The Multiple Self*. Cambridge: Cambridge University Press.

Schliesser, Eric. (2020). “Hume’s Newtonianism and Anti-Newtonianism” in *Stanford Encyclopedia of Philosophy* (online: plato.stanford.edu).

Sen, Amartya. (2017). *Collective Choice and Social Welfare: Expanded Edition.* U.K.: Penguin Books.

Sidgwick, Henry. (1981). *The Methods of Ethics, seventh edition.* Indianapolis: Hackett Publishing Company.

Simon, Herbert A. (1955). “A Behavioral Model of Rational Choice”, *The Quarterly Journal of Economics,* Vol. 69, No. 1 (February): 98-118.

Smith, Adam. (1776/1981). *An Inquiry into the Nature and Causes of the Wealth of Nations.* Indianapolis: Liberty Press.

Spiegel, Henry William. (1983). *The Growth of Economic Thought.* Durham: Duke University Press.

Strotz, R.H. (1955-1956). "Myopia and Inconsistency in Dynamic Utility Maximization", *The Review of Economic Studies,* Vol. 23, No. 3: 165-180.

Sugden, Robert. (2002). "Credible worlds: the status of theoretical models" in *Fact and Fiction in Economics: Models, Realism, and Social Construction,* Uskali Mäki, ed. Cambridge:

Thaler, Richard H. and H.M. Shefrin. (1981). "An Economic Theory of Self-Control", *Journal of Political Economy,* Vol. 89, No. 2: 392-406.

Varian, Hal R. (1984). *Microeconomic Analysis, second edition.* New York: W. W. Norton & Company.

Weithman, Paul J. (2013). *Why Political Liberalism?: On John Rawls's Political Turn*. Oxford: Oxford University Press.

Worland, Stephen T. (1973). "The Economic Significance of John Rawls' *A Theory of Justice*", in *Nebraska Journal of Economics and Business*, vol. 12, no. 4 (Autumn), pp. 119-126.

BIOGRAPHY

David C. Coker grew up an Army brat, living up and down the Eastern Seaboard. After graduating from Amherst College (English major; no economics), he had mini-careers with Starbucks Coffee and Harris Teeter Grocery before returning to school in economics, graduating from George Mason University's doctoral program in 2021. He is currently teaching as adjunct at the University of Maryland, Baltimore County. This life trajectory has tested his wife's patience, though he couldn't have done it without her.