The Erasmus Journal for Philosophy and Economics (EJPE) is a peer-reviewed bi-annual academic journal supported by the Erasmus Institute for Philosophy and Economics, Faculty of Philosophy, Erasmus University Rotterdam. EJPE publishes research on methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of interdisciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>.

EDITORS
Lennart Ackermans
Huub Brouwer
Willem van der Deijl
Erwin Dekker
Akshath Jitendranath
Vaios Kolioifotis
Marina Uzunova

ADVISORY BOARD
# TABLE OF CONTENTS

**ARTICLES**

The Persistence of the Levelling Down Objection  
*MICHAEL WEBER*  
[pp. 1-25]

Reason and Political Economy in Hume  
*ERIK W. MATSON*  
[pp. 26-51]

On the Practical Impossibility of Being Both Well-Informed and Impartial  
*SVEINUNG S. SIVERTSEN*  
[pp. 52-72]

**SPECIAL CONTRIBUTIONS**

(History of) Economic Knowledge Freed from Determinism: An Interview With *JOSEPH VOGL*  
[pp. 73-92]

Reflections on the 2018 Nobel Memorial Prize Awarded to William Nordhaus  
*J. PAUL KELLEHER*  
[pp. 93-107]

**BOOK REVIEWS**

Peter J. Boettke’s *F.A. Hayek: Economics, Political Economy, and Social Philosophy*  
*LACHEZAR GRUDEV*  
[pp. 108-113]

Tyler Cowen’s *Stubborn Attachments*  
*JOSEPH HEATH*  
[pp. 114-124]

Kate Raworth’s *Doughnut Economics*  
*ERIK SCHOKKAERT*  
[pp. 125-132]
PHD TESIS SUMMARIES

Understanding with Models
PHILIPPE VERREAUT-JULIEN [pp. 133-136]

JAMES GRAYOT [pp. 137-144]

U.S. Economics and the Quest for Scientific Authority
CAMILA OROZCO ESPINEL [pp. 145-149]
The Persistence of the Leveling Down Objection

MICHAEL WEBER
Bowling Green State University

Abstract: According to the Leveling Down Objection, some, if not all, egalitarians must concede that leveling down can make things better in a respect—in terms of equality. I argue, first, that if this is true, then it is hard for such egalitarians to avoid the even more disturbing result that leveling down can be better all-things-considered. I then consider and reject two attempts to take this particular sting out of being an egalitarian. The first is Tom Christiano’s argument that the egalitarian is not forced to concede that the leveled down state is better with respect to equality, on the grounds that if equality with respect to any X matters, e.g. welfare, then value must be assigned to X, rendering the leveled down state inferior even with respect to equality. This is insufficient, I argue, to justify his requirement that Pareto-inferior states cannot be better with respect to equality. The second, offered by both Ingmar Persson and Campbell Brown, is to argue that prioritarianism—a central rival to egalitarianism—is also subject to the Leveling Down Objection. Persson and Campbell fail, I claim, because their arguments turn on increases in measures in leveled down states that have no value on the prioritarian view.

Keywords: Leveling Down Objection, egalitarianism, prioritarianism, Tom Christiano, Pareto-efficiency, conditional egalitarianism

JEL Classification: D630

I. INTRODUCTION

The Leveling Down Objection has convinced many to reject egalitarianism. The objection, put simply, is that egalitarians must accept, implausibly, that it is an improvement, at least in some respect, if the better-off are reduced to the level of the worse-off without the worse-off becoming any better off. Egalitarians have responded to the

AUTHOR’S NOTE: For valuable feedback on earlier versions of this paper, I thank David Dick and audiences at the 2016 Pacific Division meeting of the American Philosophical Association, Wayne State University, and the 2016 meeting of the Philosophy, Politics, and Economics Society. I also received helpful feedback from anonymous referees for this journal.
Leveling Down Objection in a number of ways. One of the most interesting is Tom Christiano’s (2008) argument that the value of equality and the value of what is to be equalized, e.g. welfare, are deeply intertwined, such that it is not in any respect better if the better-off are reduced to the level of the worse-off without the worse-off becoming any better off. The resulting view appears to be a brand of what Andrew Mason (2001) calls Conditional Egalitarianism, according to which (greater) equality has final (or non-instrumental) value if and only if some are better off. Another interesting response, suggested independently by Ingmar Persson (2013) and Campbell Brown (2003), suggests that being subject to the Leveling Down Objection is not so bad because attractive alternative views such as prioritarianism are similarly subject to the Leveling Down Objection. Though both of these responses are interesting, neither is ultimately successful, or so I shall argue. A bit of background is required first, and is provided below.

It should be said, even before that, that if the egalitarian responses to the Leveling Down Objection that I discuss do indeed fail, it does not mean that egalitarianism must be rejected. First, as I will explain below, some egalitarian views—namely, those that are deontic—may be able to dismiss the Leveling Down Objection as simply inapplicable, though I cast a bit of doubt on this. Second, other egalitarians—so-called telic egalitarians—can simply bite the bullet and accept the consequence that a leveled-down state is better in a certain respect. This may be a hard bullet to bite, especially if, as I will argue, it is difficult to allow that a leveled down state is better in a respect without exposing oneself to the possibility that it is better all-things-considered. But the thrust of my argument is that recent attempts to avoid the Leveling Down Objection are inadequate, and thus biting this particular bullet may be unavoidable.

II. THE LEVELING DOWN OBJECTION

As just suggested, there are many different egalitarian views. It is common, following Derek Parfit (1997), to divide them into the two broad categories I identified above: telic (short for teleological) egalitarianism and deontic (short for deontological) egalitarianism. Parfit says that telic versions of egalitarianism are those that accept that it is
bad in itself if some people are worse off than others.\(^1\) Deontic versions of egalitarianism, in contrast, deny that it is bad in itself if some people are worse off than others. Rather, according to deontic versions of egalitarianism, inequality is unjust when it is the result of wrong-doing by some person, persons, or human institution. In such a case, as Parfit emphasizes (1997, 208), it is not the inequality itself—the state of affairs in which some are worse off than others—that it unjust; rather, it is the way in which it was produced, or came about.

Some have complained that there are egalitarian views that do not fit into either of Parfit's categories—that Parfit's categories are inadequate to capture the full range of egalitarian views. Mason (2001), for instance, argues that his Conditional Egalitarianism does not fit into either category. It doesn't fit into the teleological category because it denies that it is *always* bad in itself if some people are worse off than others. However, definitions of telic egalitarianism broader than Parfit's could encompass Mason's Conditional Egalitarianism while maintaining the key teleological focus on the value of states of affairs. Martin O'Neill (2008) complains that instrumental versions of egalitarianism are excluded. O'Neill is right, though Parfit's categories were proposed only to categorize different non-instrumental versions of egalitarianism.\(^2\)

Be that as it may, Parfit (2007) argues—and many have followed him in arguing—that at least telic egalitarianism has a troubling consequence: in at least one respect, namely, with respect to equality, an initially unequal situation can be improved by simply reducing the better-off to the level of the worse-off, as in Figure 1, in which (following a well-established convention) the width of each column represents the size of a given population or sub-population and the height represents that population’s level of welfare (or whatever one takes to be the appropriate object of equalization, e.g., resources, opportunity for welfare, or capabilities).\(^3\)

---

1. Frequently it is added that inequality is bad in itself only if it is not the product of choices people make, or is otherwise their fault. This does not bear on present concerns.
2. My interest here is limited to non-instrumental versions of egalitarianism, which have received the most philosophical attention in recent years. Instrumental egalitarianism is clearly not subject to the Leveling Down Objection, and may be the preferred egalitarian view for that reason, as O'Neill has argued.
3. The objection need not be stated in terms of one state being *transformed* into another. It can be made simply by suggesting that at least telic egalitarianism has the consequence that state of affairs B in Figure 1 is better than state A with respect to equality. So too with the examples to follow.
The Leveling Down Objection is the suggestion that this is an unacceptable consequence. A common way to support the Leveling Down Objection is to appeal to a principle according to which one situation or state of affairs cannot be better than another in any respect if no one is better off (and similarly cannot be worse in any respect if no one is worse off), though some think that a weaker principle is sufficient.\(^5\)

There are less extreme cases of leveling down that seem no less troubling. For instance, inequality can be reduced by lowering the level of the better-off some, but not enough to lower them to the level of the worse-off, as in Figure 2.

---

\(^4\) To avoid complications associated with Parfit's Non-Identity Problem, it is to be assumed that it is the same set of people in states A and B.

\(^5\) Parfit (1997) calls the wider principle the person-affecting principle; Larry Temkin (2002) calls it *The Slogan*. Temkin (1987, 1993a, 2002, 2003a, 2003b) has argued that we should reject *The Slogan*. However, this is controversial, and engaging in the controversy is beyond the scope of this paper. Nils Holtug (2007b), among others, argues that the Leveling Down Objection does not rest on *The Slogan*, but rather on a narrower and more defensible relative of it. Roger Crisp (2003a, 2003b) argues similarly. As such, the Leveling Down Objection cannot be simply dismissed by casting doubt on *The Slogan*. 
It is also leveling down if the levels of both the better-off and the worse-off are reduced, but the level of the better-off is reduced more. To be an instance of leveling down, all that is required is that inequality is reduced in virtue of some being made worse off without any others being made better off.⁶

Some, including Parfit (1997), think that deontic egalitarianism is immune to the Leveling Down Objection because it denies that inequality is bad in itself, and thus is not committed to saying that less unequal states are at least in one way better than more unequal states. Though I will not go into the matter in much detail here, it seems that nonetheless deontic egalitarianism is subject to an analogous objection. For if there is an unjust inequality, then the deontic egalitarian seems committed to the claim that rectifying the injustice is a moral improvement, even though no one is made better off. Some do not find this as paradoxical as the claim that one state of affairs can be better than another even if no one is better off. There is considerable controversy about this matter, and I do not propose to settle it here.⁷

---

⁶ Temkin (1993b) has taught us that it is often very difficult to determine when inequality has been increased or reduced, because it is a complex notion. So it may often be difficult to determine in a given instance whether it is a case of leveling down. The cases I employ, however, do not give rise to such complications.

⁷ Kasper Lippert-Rasmussen (2007) and Martin O'Neill (2008) both argue that deontic egalitarianism is equally subject to the Leveling Down Objection (or some analogous objection). However, Michael Otsuka and Alex Voorhoeve (2009, 183–184) advance a deontic egalitarian view grounded in the contractarian idea that allocations must be justifiable to each person taken separately that they claim is immune to the Leveling Down Objection.
or close analogues, then it falls outside the scope of my present concerns.

In large part because of the Leveling Down Objection, a lot of people with egalitarian (and teleological) leanings have adopted, instead, what Parfit (1997) has dubbed the Priority View, or prioritarianism. Prioritarianism is the view that benefitting people matters more the worse off people are in absolute terms. This is held to be a superior view because it a) tends toward equality in virtue of the priority it gives to benefiting those who are worse off (in absolute terms), and b) is immune to the Leveling Down Objection because, as with (orthodox) utilitarianism, the only way to improve a situation is to make someone better off.

Some egalitarians, however, have stood their ground. Larry Temkin (2002), for instance, argues that while the leveled down state may be better with respect to equality, it does not follow that (even telic) egalitarians are committed to the claim that the leveled down state is better all-things-considered, because egalitarians can be pluralists, assigning final (or non-instrumental) value to both equality and welfare. So long as the loss in terms of welfare outweighs (or otherwise defeats) the gain in terms of equality, the leveled down state will be worse all-things-considered. Therefore, egalitarians can be what Parfit calls Moderate Egalitarians, maintaining that “though it would always be in one way better if there ceased to be inequality between different people [or inequality was reduced], this outcome would be worse all-things-considered if it would be worse for some people and better for no one” (Parfit 2012, 399). Moderate egalitarians, then, concede only that the leveled down state of affairs is better in a respect, and claim that this is not so bad.

Parfit disagrees, arguing that it is more than a little troubling if egalitarianism has the result that a leveled down state is better in a respect, even if it is not better all-things-considered. It is, he says, a significant objection, even if not a decisive objection (401). His argument is that it is not enough when evaluating a moral theory to consider only what it deems better or worse, or right or wrong, all-things-considered. He insists that “we should reject moral theories which make implausible

---

8 As Parfit notes, and others such as Andrew Williams (2012) have emphasized, deontic versions of prioritarianism are possible, though most of its well-known advocates defend teleological versions.

9 Temkin (1987, 1993a, 2002, 2003a, 2003b) has offered a variety of other responses to the Leveling Down Objection, including, as noted earlier, arguing that we should reject The Slogan.

10 I am grateful to David Dick for emphasizing this point to me.
claims about which are the facts that can make outcomes better or worse, or can make acts right or wrong” (400). And while it is not absurd, Parfit thinks it is disturbing to accept that if everyone became equally badly off, that would make an outcome in some way better.\footnote{Against deontic egalitarians, it could be put this way: it is disturbing, if not absurd, to think that there can be a moral improvement if everyone became equally badly off.} As such, he concludes, the Leveling Down Objection has “considerable force” (401). Roger Crisp (2003a) and Mason (2001, 252), among others, agree.

It seems to me, however, that Parfit may underestimate the problem with Moderate Egalitarianism. According to Moderate Egalitarianism, a leveled down state can be better in some respect but it cannot be better all-things-considered. But this requires that the loss in terms of welfare is always more significant morally than whatever gains there are in terms of equality. Surely the Moderate Egalitarian must provide a justification for assigning this kind of priority to welfare over equality (and not in an ad hoc way).\footnote{Parfit (1997) suggests one reason for thinking that in leveling down cases the losses in terms of welfare will outweigh or otherwise defeat the gains in terms of equality. The idea, roughly, is that for any large increase in equality in a leveling down case there must also be a large decrease in welfare. I am not convinced by this for two reasons, however. First, as indicated previously, Temkin (1993b) has shown that equality is quite complex. As such, there is no guarantee that large increases in equality necessarily require large decreases in welfare. Second, even if it is true that large gains in equality require large losses in welfare, this does not guarantee that disvalue of the losses in terms of welfare is greater than the value of the gains in terms of equality.} Barring that, it seems that it must be conceded that if a leveled down state can be better in a certain respect, then it is at least possible that it is better all-things-considered. As such, Moderate Egalitarianism appears to be an unstable position.\footnote{This has also been noted by Dennis McKerlie (1996) and Bertil Tungodden (2003, 10).}

In fact, it can be argued that things are even worse than that for Moderate Egalitarianism. Temkin (2002, 2003a, 2003b) has argued that there are a number of values that, like equality, entail that leveling down can bring about an improvement in some respect.\footnote{Temkin refers to such values as \textit{impersonal} values.} Imagine, for instance, that on some relevant metric A is more deserving than B (in the extreme, A is a saint and B is a sinner), yet B is better off. Plausibly, it would be better with respect to desert if B were worse off, if, for instance, he were leveled down so that he is worse off than A. If Temkin is right that there are many values other than equality that are better realized through leveling down—and I think he is—then leveling down can bring about an improvement in a variety of respects, in which case it
seems all the more difficult to be absolutely confident that leveling down cannot be an improvement all-things-considered. Moderate Egalitarianism requires an argument that what is lost in leveling down is always morally more important than what is gained. This is a tall order.\textsuperscript{15}

\section*{III. Christiano's Defense of Egalitarianism}

If what I have said about Moderate Egalitarianism is true, then (telic) egalitarians cannot happily allow that the leveled down state can be better in a certain respect, as this exposes them to the possibility that the state is better all-things-considered. This is what makes Tom Christiano's response to the Leveling Down Objection in \textit{The Constitution of Equality: Democratic Authority and its Limits} so interesting, as he argues that leveling down does not even make things better in some respect—not even with respect to equality.

Christiano's case for his view that leveling down is not an improvement from an egalitarian point of view has two major components. First, he argues that the value of welfare (or whatever else one takes to be the proper object of equalization) is “built into” the principle of equality (Christiano 2008, 33). There is, he says, “an internal connection between the rationale for equality and the value of the relevant fundamental good that is equalized” (33). He argues for this connection with a compelling example. Imagine that there is enough bread to give every person more bread than he or she could possibly use (in which case the bread that one did not eat or otherwise use would have no trade-value). There is, then, a surplus of bread after each person has been given as much as he or she could possibly use. Christiano claims, quite plausibly, that it simply does not matter how this surplus is distributed, in particular whether it is distributed equally or not (or not distributed at all). For the equal distribution of the bread only matters, if it matters at all, if the bread has value to the potential recipients. More generally, it is only if more X is better than less that equality with respect to X matters.\textsuperscript{16}

\textsuperscript{15} Indeed, one of the main points of this paper is to show just what a tall order it is. See sections III and IV.

\textsuperscript{16} Christiano misses that with things of negative value, e.g., pain or punishment, equality can matter even though more is not better than less. He could unproblematically modify the view to be ‘it is only if more X is better than less, or less better than more, that equality with respect to X matters’.
that more is better than less” (34). Thus, if equality with respect to X matters, it must be that more X is better than less.\textsuperscript{17} Thus, Christiano claims, an egalitarian (with respect to welfare) cannot be indifferent between two perfectly equal states when one has members at a higher level of welfare, as in the Figure 3 below.\textsuperscript{18}

![Figure 3. Perfectly equal states with different levels of welfare](image)

The egalitarian must prefer state of affairs A, according to Christiano, because while there is no difference between A and B in terms of the equality of the distribution of welfare, A has greater (total and average) welfare. In a simplified two-person world, then, the egalitarian must prefer a world in which both people are at a welfare level of five (5,5) to a world in which both people are at a welfare level of 3 (3,3), as in Figure 4.

<table>
<thead>
<tr>
<th>Person</th>
<th>State of affairs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
</tr>
</tbody>
</table>

\textbf{Figure 4.}

\textsuperscript{17} Some positive values are not such that more is better than less. Friendship, for example, is valuable, but it does not follow that one should try to maximize the number of friends one has (though perhaps one should maximize the quality of the friendships one does have). On Christiano’s view, equality with respect to such values cannot matter (because more is not better than less). This might be thought to be a dubious consequence.

\textsuperscript{18} This is significant because egalitarians tend to think that egalitarianism itself is indifferent between these states; the superiority of A comes only with a pluralistic egalitarianism that independently assigns final value to welfare (or whatever one takes to be the appropriate object of equalization).
The second component of Christiano’s defense of egalitarianism against the Leveling Down Objection starts by noting that all an egalitarian has to claim about unequal states is that in any such state there is something amiss or flawed, that things are less than ideal, and, as such, could be better in terms of equality. This can be done, however, in such a way that it will not be the case that any more equal state of affairs is better in terms of equality. Central to this second component of the defense is the idea that that we often explain how something is amiss or less than ideal in some respect by comparing it to an idealized version of itself. For instance, a house-hunter with a growing family might say that a four-bedroom house that he has been shown is not quite perfect because there are two bedrooms on each floor rather than all four on the same floor.\(^{19}\) It would be perfect, he might say, if it were exactly the same, except for the fact that all the bedrooms were on the same floor.\(^{19}\) Similarly, then, any unequal state of affairs can be shown to be flawed, or less than perfect in terms of equality, by comparing it to a more idealized version of itself. For instance, consider the simplified two-person world or state of affairs in which one person, A, has a welfare level of seven, and the other, B, has a welfare level of three (7,3), as in S3 in Figure 5.

<table>
<thead>
<tr>
<th>Person</th>
<th>State of affairs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 5.**

This state of affairs (S3) can be shown to be flawed or imperfect from an egalitarian point of view by comparing it to an idealized version of itself. The idealized version of itself, Christiano argues, “is to be understood as an equal and Pareto optimal distribution of the goods that are present in the best Pareto optimal state” (38). Thus, he claims, the ideal relative to S3, where the distribution is (7,3), is actually S1 (5,5),

\(^{19}\) Of course a preference for all bedrooms being on the same floor is not universal. I am just assuming that the house-hunter has this preference for illustrative purposes.
a perfectly equal state which preserves the (total and average) welfare in the original unequal state. The reason for preserving in the ideal the (total and average) welfare in the original unequal state comes from the first component of Christiano’s defense of egalitarianism, namely that (the value of) welfare is ‘built into’ equality, such that more welfare is better than less. Relative to the ideal S1 (5,5), then, S3 (7,3) is flawed or imperfect from an egalitarian point of view, just as the actual house the house-hunter viewed is flawed or imperfect relative to the idealized version of itself in which all the bedrooms are on the same floor.

Now imagine that the house-hunter's real estate agent were to show him a three-bedroom house with all the bedrooms on the same floor. Given that the house-hunter has a growing family, it is easy to imagine that he would prefer the four-bedroom house with two bedrooms on each floor. For, given his growing family, the four-bedroom house with two bedrooms on each floor is closer to the ideal than the three-bedroom house with all the bedrooms on the same floor. We might say, then, that the while the four-bedroom house with two bedrooms on each floor is not perfect because the bedrooms are on separate floors, it does not follow from this that any house with all the bedrooms on the same floor is better.

There might be some confusion here, because S1 (5,5) is not Pareto-optimal relative to S3 (7,3); it is, rather, Pareto-incomparable (though it is more equal according to the Pigou-Dalton Principle). However, Christiano is not claiming that the ideal is Pareto-optimal relative to S3 (7,3). Rather, he is claiming that the ideal is an equal and Pareto-optimal distribution of the total units of welfare (10) in S3. The Pareto-optimality requirement here rules out equal distributions in which less than all ten units are distributed, e.g., (4,4), (2,2), and (0,0).

It might be argued that in the real world it is not possible to preserve the welfare in the unequal state (7,3) in a move to a more equal state because there are always transactions costs in redistribution. (5,5), then, is not a feasible and more equal alternative to (7,3). Feasible and more equal alternatives will have less total welfare, e.g., (4.8, 4.8), or (5, 4.7). Although he does not consider this particular concern, Christiano does say that the ideal equality point need not be feasible (38). This raises two concerns. First, if the ideal need not be feasible, why not make (7,7) (or even (14,14)) the ideal of (7,3)? Second, it might seem inappropriate to evaluate states relative to an impossible ideal. If (4.8, 4.8) is the best equal and feasible alternative to (7,3), it seems odd to deem it unjust because it falls short of an unrealistic (5,5). Christiano has replies to each of these concerns. To the first, he says that the ideal should not be "purely utopian" (38). To the second, he argues that "justice is not always feasible as when two equally deserving persons create goods that are unequal in value and cannot be divided up" (38-39). These two replies together constitute what might be called 'moderate idealism', according to which it is acceptable to idealize—to abstract away from reality—only to a limited extent. Such moderate idealism, which avoids the extremes of realism and utopianism, is the emerging norm in ethics and political philosophy (see Vallier and Weber 2017). So perhaps Christiano can adequately address these concerns.
So too, Christiano claims, with the unequal state S3 (7,3). It is flawed, or imperfect, from the egalitarian point of view because of its shortcomings with respect to S1 (5,5). But it does not follow from this that any more equal state such as S2 (3,3) is better in terms of equality. For just as the four-bedroom house with bedrooms on different floors is closer to the ideal than the three-bedroom house with all the bedrooms on the same floor, Christiano claims, S3 (7,3) is closer to the ideal, S1 (5,5), than is S2 (3,3). Thus, S2 is not better than S3 in terms of equality. Of course this claim—that the unequal state S3 (7,3) is closer to the ideal S1 (5,5) than the leveled down state S2 (3,3)—requires a rule or set of criteria for judging closeness to the ideal—for determining when one state is closer to, or a better approximation of, the ideal. Christiano does not actually provide such a rule or set of criteria. Instead, he suggests constraints on any such rule or set of criteria. The most important of these, for present purposes, is a Pareto-constraint: “the rule must never favor Pareto-inferior states over Pareto-superior states” (39). With this constraint, S3 (7,3) is closer to the ideal, S1 (5,5), than the leveled down state S2 (3,3) because it is Pareto-superior. And the same will be true with any leveled down state, e.g., (2,2), (5,3), and (5,2). Thus, Christiano concludes, leveled down states are not better in egalitarian terms.

IV. OBJECTIONS TO CHRISTIANO’S DEFENSE

There are, however, several objections to Christiano’s defense of egalitarianism against the Leveling Down Objection. In the first part of the defense, he argues that welfare concerns are ‘built into’ equality of welfare. He is right, I will grant, that egalitarians (about welfare) must agree that more welfare is better than less, and must, therefore, prefer S1 (5,5) to S2 (3,3). But this does not make S2 better in egalitarian terms. For the argument shows only that the egalitarian is committed to preferring more welfare than less, and thus preferring S1 to S2 all-things-considered. A commitment to equality of X may entail preferring

---

22 Note that (5,3) and (5,2) are examples of the less extreme leveling down identified in Section II.
23 In a co-authored paper, Christiano and Will Braynen (2008) present the same basic argument that Christiano offers in his book. What they add (409–419) is some detailed discussion of alternative ways of determining how far a given unequal state is from the (corresponding) ideal egalitarian distribution. They argue for a divergence rule of approximation over a difference summing rule of approximation. The difference between the two rules does not bear on my concerns here. What does matter is that both assume a Pareto-constraint: Pareto-superior states are always closer to the ideal than Pareto-inferior states.
more X to less; however, it does not show that X is a part of—‘built into’—equality. There is a difference between egalitarians being committed to the value of welfare and welfare being literally built into or part of equality. Consider a different case. Some dance schools require their dancers to take ballet classes, even if the student is only interested in another form of dance, e.g., jazz or modern. The reason for this is that many of the steps, turns and progressions in jazz and modern dance just are ballet movements. In this case, it seems to me, ballet is a part of jazz or modern dance; it is built in, literally. But this does not seem to be the relation between equality and welfare, even if Christiano is right about what egalitarians are committed to. For in the case of equality, on Christiano’s view, one is committed to something conceptually distinct. Thus, it seems to me, S1 (5,5) is not better than S2 (3,3) in egalitarian terms. Rather, it is better only in terms the egalitarian is committed to.

Now consider the second component of Christiano’s defense of egalitarianism against the Leveling Down Objection. I will grant that what he says about establishing the ideal is correct—and, thus, that relative to S3 (7,3), S1 (5,5) is the ideal. What seems dubious, instead, is what he says about approximations to the ideal. Perhaps it is intuitive that S3 (7,3) is closer to the ideal than S2 (3,3). But, as noted earlier, there are less extreme cases of leveling down, for instance when the welfare level of the better-off is reduced, but not so much as to make them equal to the worse-off. So, for instance, a state S4, in which the better-off have a welfare level of 5 while the worse-off are at 3 (5,3), is a leveled down state relative to S3 (7,3). See Figure 6, which adds S4 to Figure 5.

<table>
<thead>
<tr>
<th>Person</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 6.

It seems to me that it is far from intuitive that S3 (7,3) is closer to the ideal (5,5) than S4 (5,3). To insist that S3 (7,3) is closer to the ideal...
seems to attach greater significance to (preserving) welfare than to having a more equal distribution of welfare, because the more equal distribution in S4 (5,3) does not count as much in determining which state is closer to the ideal as does the lower level of welfare. Indeed, Christiano's Pareto-constraint seems to assign lexical priority to (preserving) welfare over realizing a more equal distribution of that welfare. But even if it is agreed that in some important sense the egalitarian (with respect to welfare) is committed to placing value on welfare, this does not justify attaching greater significance to welfare such that S3 (7,3) is closer to the ideal (5,5) than S4 (5,3), or even S2 (3,3). Perhaps it is enough to show that egalitarians must balance considerations of equality (of welfare) and considerations of total or average welfare. But, depending on how the balance is struck, S4 (5,3) may be a better approximation of the ideal (5,5) than S3 (7,3). Indeed, S2 could be a better approximation.

Of course Christiano might dismiss such appeals to intuition and alternative weightings of equality and welfare, insisting that he has a principled reason for thinking that S3 (7,3) is closer to the ideal than S2 (3,3), or any other leveled down state: the rules for approximation to the ideal must favor Pareto-superior states. However, simply insisting on such a Pareto-constraint would beg the question, as the debate about the Leveling Down Objection just is a debate over whether values must be Pareto-compatible (because leveled-down states are Pareto-inferior). Christiano does not merely insist on the Pareto-constraint. He says that it “follows from the importance of the well-being of all persons that is essential to the principle of equality” (Christiano 2008, 40). But as I have argued, even if it is true that a concern for equality of welfare commits one to valuing more welfare than less, it does not justify placing concerns for welfare ahead of equality. If one cares at all about equality of X, then surely one is committed to the claim that more (total or average) X is better than less other things equal. But in cases of leveling down, other things are not equal: there is less welfare, but what there is is more equally distributed. As such, the Pareto-constraint simply formalizes assigning greater significance or value to (preserving) welfare

---

24 Perhaps there is an intuitive sense that S3 (7,3) is closer to S1 (5,5) than S4 (5,3) is, because S3 and S1 have the same sum and average utility, but differ in terms of equality, while S4 and S1 have a different sum and average utility and differ in terms of equality. But the point still holds, because saying that all-things-considered S3 is closer to S1 than S4 requires assigning greater significance to utility considerations that equality considerations, because though both S4 and S3 are not as equal as S1, S4 is closer to S1 in terms of equality than S3 is.
than the (more) equal distribution of welfare. It simply formalizes the unfounded assumption that improvements cannot be made by trading off welfare for greater equality.

V. CONDITIONAL Egalitarianism

Although he never makes the suggestion explicitly, Christiano’s view is plausibly viewed as a version of Conditional Egalitarianism.\(^{25}\) According to Mason, recall, Conditional Egalitarianism is the view that (greater) equality is non-instrumentally valuable if and only if some are better off.\(^{26}\) Christiano’s argument ultimately reaches exactly this conclusion.\(^{27}\) For he argues, in short, that a state of affairs X is better in terms of equality only if it is closer to the relevant egalitarian ideal, and Pareto-inferior states can never be closer to the ideal (than Pareto-superior states). It follows from this that leveled down states cannot be better with respect to equality, because, as noted earlier, leveled down states are by definition Pareto-inferior.\(^{28}\)

Mason supports his claim that the value of (greater) equality is conditional on some being better off by appeal to intuition (Mason 2001, 250–251). But this can seem unsatisfactory. If (greater) equality is non-instrumentally valuable in some situations—when some are better off—and not in others, then there should be some explanation for this variability.\(^{29}\) If someone suggests that chocolate cake (or the pleasure of eating chocolate cake) is good only if there is a full moon, we would surely insist on some explanation for this. What does the full moon have to do with chocolate cake? Of course this is a rhetorical question: the

\(^{25}\) Thanks again to David Dick for helping me to appreciate this point.

\(^{26}\) Holtug (2007a, 50–52) argues compellingly that Conditional Egalitarianism should require further that it is more equal because some are better off. Temkin (1987, 171) makes the same point when he suggests a “person-affecting version of egalitarianism” according to which “concern about inequality is not merely a concern that inequality be removed, but that it is removed in a certain way, so as to benefit the worse off”.

\(^{27}\) Bertil Tungodden and Peter Vallentyne (2005) also reach a similar conclusion.

\(^{28}\) Weaker forms of Conditional Egalitarianism are possible. Dale Dorsey (2013, 56) suggests the view that equality has non-instrumental value only if it does not make the worse-off any worse-off. However, as Dorsey admits, such Conditional Egalitarianism accepts that in some instances a leveled down state is better (at least with respect to equality), in particular when the very well off (or the ‘ultra-rich’) are leveled down (and those worse off neither gain nor lose). Perhaps, as Dorsey suggests, this degree of leveling down is acceptable. In this paper, however, I am restricting myself to considering whether there is an egalitarian view that can completely avoid the Leveling Down Objection. Dorsey freely admits that his version of Conditional Egalitarianism does not.

\(^{29}\) Holtug (2007a, 56; 2010, 193–198) has emphasized this point. Also see Segall (2016, 29–31).
full moon has nothing to do with chocolate cake.\textsuperscript{30} And this is why we would not accept that the value of chocolate cake is conditional on there being a full moon. So if we are to accept that the value of equality is conditional on some being better off, there must be a connection between equality and some being better off sufficient to justify making the non-instrumental value of (greater) equality conditional on some being better off.

As we have seen, Christiano does argue for a connection between equality and welfare. He thinks that if equality with respect to X matters only if more X is better than less (recall the bread example). Thus, if equality with respect to welfare matters, then it must be that more welfare is better than less.\textsuperscript{31} This is insufficient, however, for reasons identified in the previous section. For Christiano establishes only that egalitarians (with respect to welfare) cannot be indifferent to welfare levels. But this is a far cry from a requirement that equality has value only if some are better off (just as it is a far cry from grounding a Pareto-constraint on rules of approximation to the ideal). If one accepts that the egalitarian with respect to welfare cannot be indifferent to welfare levels, then such an egalitarian must accept that higher welfare levels are better \textit{other things equal}. But, as indicated in the previous section, with leveling down other things are not equal: in the leveled down state, the distribution of welfare is more equal. What was said in the previous section, then, shows that Christiano does not have an adequate explanation for holding that the value of equality is conditional on some being better off (just as he has no ground for a Pareto-constraint on rules of approximation to the ideal). Rather, he establishes only that the egalitarian (with respect to welfare) cannot be indifferent to welfare levels. But this is compatible with a leveled down state being better in virtue of the gain in equality being morally more significant than the loss in terms of welfare.\textsuperscript{32}

\begin{footnotes}
\footnotetext[30]{Of course there could be some kind of contingent connection, e.g., the best baker of chocolate cake only bakes when the moon is full. But typically there is not.}
\footnotetext[31]{Of course few, if any, would deny that more welfare is better than less (other things equal). But what matters for Christiano is that such a view is \textit{internal} to egalitarianism, rather than a separate, independent commitment.}
\footnotetext[32]{One might, alternatively, try to ground Conditional Egalitarianism by directly appealing to the person-affecting principle (\textit{The Slogan}, as Temkin calls it): something is good, e.g., greater equality, only if it is good for someone; thus, greater equality is good only if some are better off. Two points can be made here. First, Christiano assiduously avoids appeal to the person-affecting principle. So this is not his approach. The likely reason he avoids it is the second point: the (justification of the) person-affecting principle is a well-known hornet’s nest. Additionally, Holtug (2007a, 57-58)}
\end{footnotes}
Mason (2001) thinks that the demand for an explanation for the conditional value of equality is unreasonable—that in fact no explanation is required for the value of (greater) equality being conditional on some being better off. Many widely-accepted views about non-instrumental value, he notes, are defended by appeal to intuition, e.g., the claim that pleasure has non-instrumental value only if it is the product of a morally permissible act (or contemplation of such an act); pleasure that results from the performance or contemplation of an immoral act has no non-instrumental value. Indeed, he suggests that no other defense could be given. He says this explicitly about the case of the conditional value of equality—that it can only be defended by appeal to intuition (251).

It is hard to know how to settle such a dispute about whether an explanation is needed for these kind of claims about non-instrumental value. Even if it is agreed that justification has to end somewhere, people can disagree about where it ends—about when bedrock has been reached. With respect to the claim that the value of pleasure is conditional, Nils Holtug argues that to the extent that we are inclined to think so it is because there is an explanation for it, namely that pleasure is an inappropriate attitude to take towards immoral actions (Holtug 2007a, 61). Whether we agree with Holtug’s explanation or not, the fact that such an explanation can be proposed undermines Mason’s claim that such claims about non-instrumental value can only be defended by appeal to intuition—that it is somehow inconceivable that a deeper explanation can be provided. Perhaps there is a difference between the conditional value of pleasure and the conditional value of equality, such that the there is more reason to think that no further explanation can even be proposed in latter case. But there is surely no obvious difference, which puts the burden of proof on the side of those who think no further explanation is required (because somehow further explanation is inconceivable).

It might be held that no further explanation is required in cases like those under discussion if a view is widely accepted in the absence of an explanation—when those who accept it do so without being able to
explain why they do. The thought here is that this is a hallmark of having reached bedrock: the best explanation for wide acceptance without an explanation is that there is no further explanation (possible). Some might make an argument along these lines for the conditional value of pleasure: it requires no further explanation because it is widely-accepted even in the absence of an explanation. But there are at least two problems with this approach, even if we ignore the fact that the conditional value of pleasure is disputed. First, it is unclear that the best explanation for wide acceptance in the absence of an explanation is that bedrock has been reached, because such agreement might be the product of an underlying but not fully appreciated explanation. Indeed, I think this is Holtug’s point about the conditional value of pleasure: to the extent that we accept the conditional value of pleasure it is because of an inchoate sense that pleasure is an inappropriate attitude to take towards immoral actions (or an inchoate sense of some other explanation). Second, even if it is allowed that there is widespread acceptance of the conditional value of pleasure, the conditional value of equality is surely quite contentious. Such disagreement undermines the suggestion that no further explanation is required because bedrock has been reached. It is further undermined by the fact that in the case of the conditional value of equality the dissenters are non-partisan: there are both egalitarians, e.g., Moderate Egalitarians such as Temkin, and critics of egalitarianism, e.g., Parfit, that think that a leveled down state is better with respect to equality—that the value of equality is not conditional on some being better off. These dissenters are divided on a different matter, as indicated earlier: they disagree about the significance of the fact that on the (telic) egalitarian view a leveled down state is better in a certain respect. Dissenters can sometimes be dismissed on the grounds that their dissent is ideologically driven, though people are often inclined to do so much too readily, without sufficient justification. But because dissent in the case of the conditional value of equality is non-partisan, such a dismissal is a non-starter.

Neither Christiano nor Mason, then, offers an adequate defense of Conditional Egalitarianism. It requires a more robust defense, one that provides a satisfactory explanation for the conditional value of equality.

---

33 The best examples of this come from sports-talk radio, where a caller’s critical opinion of a team, or a player on a team, are immediately dismissed when it is revealed that the caller is a fan of a rival team.
VI. THE PRIORITARIAN ALTERNATIVE

Recall that those with egalitarian (and teleological) leanings have an alternative to (telic) egalitarianism, namely prioritarianism. It has strong egalitarian tendencies in virtue of the priority it assigns to benefitting those worse off in absolute terms, and seems immune to the Leveling Down Objection because according to prioritarianism the only way to improve a situation is to make someone better off. As such, it presents a challenge to (telic) egalitarianism, as it offers most of the benefits without the costs.\(^{34}\) As noted earlier, Ingmar Persson (2013) and Campbell Brown (2003) both suggest an interesting response to this challenge: they argue that, contrary to initial appearances, prioritarianism too is subject to the Leveling Down Objection. Persson argues that prioritarianism is subject to the Leveling Down Objection because according to prioritarianism the average moral value of benefits (or utility) is higher in the leveled down state because benefits lost by those who are better off have relatively less moral value:

\[
\text{In the leveling down situation [. . .] there is a reduction of the sum of benefits since the better off lose something, and the worse off gain nothing. But the average moral value of benefits goes up because it is the benefits with the lowest such value that are removed, i.e., the benefits which make some better off than others. (Persson 2013, 4)}
\]

The leveled down state, then, is better with respect to the average moral value of benefits. It is better in some respect, then, even though no one is better off.

Brown's case that prioritarianism is subject to the Leveling Down Objection is rooted in his analysis of what it is for something to be better 'in a respect’. To simplify a bit, according to Brown's analysis A is better than B in respect r if and only if were there no other relevant differences between A and B the difference between A and B in respect r would suffice to make A better than B all-things-considered. When this analysis is applied to a standard case of leveling down, as in Figure 1, Brown claims, the prioritarian “is committed to the claim that B is better than A, in respect of equality” (Brown 2003, 117).\(^{35}\)

\(^{34}\) Temkin (2002) complains that it only offers some of the benefits—that some egalitarian intuitions cannot be captured by prioritarianism.

\(^{35}\) Shlomi Segall (2015) also argues that prioritarianism cannot establish its superiority to egalitarianism by appealing to the Leveling Down Objection, though on slightly different grounds. His argument is that to avoid a troubling implication of prioritarianism in the case of an intrapersonal dilemma proposed by Otsuka and
At first pass, it is not at all clear either Persson or Brown has shown that prioritarianism is on a par with egalitarianism with respect to the Leveling Down Objection. Persson is right that in the leveled down state the average moral value of benefits is higher. But this does not seem to matter, because prioritarianism does not assign final (or non-instrumental) value to the average moral value of benefits. Just as the average moral value of benefits is higher in the leveled down state, average height could be greater, as could the percentage of the population that likes scrambled eggs. But it is no embarrassment to prioritarianism if average height is greater in the leveled down state, or if the percentage of the population that likes scrambled eggs is greater, because these are not things that have final (or non-instrumental) value on the prioritarian view. Similarly, then, it is no objection to prioritarianism that in the leveled down state the average moral value of benefits is greater because what matters to prioritarianism—what has final value—is not the average value of benefits but the total value of benefits. The average value of benefits may be greater, but that does not make it in any respect better if the average moral value of benefits has no value.

Another way to put it is this. The problem for egalitarianism is that the leveled down state is better in egalitarian terms. If prioritarianism is on a par with egalitarianism with respect to the Leveling Down Objection, then it must be the case that the leveled down state is better in prioritarian terms. But the fact that in the leveled down state the average moral value of benefits is higher does not establish that it is

Voorhoeve (2009), prioritarianism must reject the Principle of Personal Good (PPG), a principle that Segall claims must be accepted to ground the Leveling Down Objection to egalitarianism. Though he does not make the point explicitly, rejection of PPG does seem to make one vulnerable to the Leveling Down Objection (or some analogous objection). It is beyond the scope of this paper to consider Segall’s very interesting and distinct argument, in part because, as suggested, it flows from consideration of the very detailed critique of prioritarianism leveled by Otsuka and Voorhoeve. I discuss Otsuka and Voorhoeve’s argument, and prominent responses to it, in Weber (2014).

Porter (2011, 200–201) circles close to this point, as I note in the following.

While average height or the percentage of population that likes scrambled eggs might be greater in the leveled down state, the average moral value of benefits must be greater in the leveled down state. However, it seems to me that even with this stronger modal connection prioritarianism is not committed to assigning final value to the average moral value of benefits.

The point here, of course, is that ‘greater’ is a descriptive term, while ‘better’ is evaluative. Perhaps sometimes we use ‘better’ in a purely descriptive sense. Be that as it may, Persson’s argument requires that the leveled down state is better in the evaluative sense.
better in prioritarian terms (because prioritarianism does not assign final value to the average moral value of benefits). Persson, therefore, has not shown that prioritarianism is on a par with egalitarianism with respect to the Leveling Down Objection.

If this is right about Persson’s case, then it seems the same can be said about Brown’s case that prioritarianism is equally subject to the Leveling Down Objection. On Brown’s analysis, the prioritarian is committed to saying that the leveled down state is better in terms of equality. But equality does not have final value according to prioritarianism. So the leveled down state is not better in prioritarian terms. There may be more equality in the leveled down state, but that does not make it better in virtue of the increased equality if equality has no final value.39

Although he does not develop the point in the same kind of detail, Thomas Porter has suggested a similar argument against Persson’s argument that prioritarianism is subject to the Leveling Down Objection.40 In cases of leveling down, as Persson has emphasized, the average moral value of benefits (or utility) increases despite a decrease in overall benefits (or utility). Porter asks skeptically: “When the total and average of some value that you care about go in different directions, why not focus on the total alone and assign no moral significance to the [...][increase] in the average?” (Porter 2011, 201).

Persson (2011), however, has argued in reply to Porter that it is not an option for the prioritarian to hold that the average moral value of benefits (utility) is a matter of indifference. He argues that if no value at all is assigned to it, then the view is indistinguishable from utilitarianism. The thought is this. For utilitarianism, all units of benefit (utility) have the same value. It does not matter who has the benefit, nor does it matter what his or her benefit level is. Therefore, utilitarianism entails that the average moral value of benefits is invariant. This is not true for prioritarianism, because, according to prioritarianism, some units of benefit have less (moral) value than others. In particular, units of benefit at a higher level of total benefit are less morally valuable than those at a lower level. Thus, if a person at a low total benefit level of 3 units gains an additional unit, this is more morally valuable than if a person at a high benefit level of 10 gains an additional unit. Because on

39 So, the point is, ‘more’ is descriptive just as ‘greater’ is.
40 Holtug (2010, Chapter 8) discusses this issue at some length, and reaches similar conclusions.
a prioritarian view the value of additional units of benefit (utility) decreases, the average moral value of benefits decreases with Pareto-improvements and increases with leveling down to a Pareto-inferior state. Imagine, for instance, a two-person world $W_1$ in which person A has two units of benefit and person B has just one (2,1). Assume a prioritarian view according to which the first unit of benefit (utility) for any person has a moral value of 10 while the second has a moral value of 4. In $W_1$, then, the total moral value of the benefit is 24, and the average moral value of a unit of benefit is 8. Now imagine that B is provided with another unit of benefit (utility), resulting in the two-person world $W_2$ in which both A and B enjoy two units of benefit or utility (2,2). In $W_2$, the total moral value of the benefits is higher, at 28. However, the average moral value of benefits has decreased to 7. Now consider a different variation from $W_1$: A loses a unit of benefit (utility), so that now both A and B have just one unit of benefit or utility (1,1). Call the resulting state $W_3$. In $W_3$, the total moral value of benefit has decreased from 24 to 20, while the average moral value of units of benefit or utility has increased from 8 to 10. What distinguishes prioritarianism from utilitarianism, then, is that the average moral value of benefits is variable. As such, Persson suggests, if one denies the moral significance of the average value of moral benefits, then one has abandoned prioritarianism in favor of utilitarianism.

There is, however, a simple reason to disagree with Persson’s reply: the prioritarian is not denying that the average value of benefit varies; rather, the prioritarian is simply claiming that the (variable) average value of benefit (in a given state of affairs) has no moral significance or value in itself. Insofar as prioritarianism maintains that the average moral value of benefit is variable, it is not equivalent to utilitarianism. Therefore, neither Persson nor Brown seems to have successfully shown that prioritarianism is equally vulnerable to the Leveling Down

---

41 Porter concludes similarly: “Indifference to changes in average prioritarian unit value need not imply indifference to changes in actual prioritarian value, since there can be changes in average prioritarian unit value without any changes in actual prioritarian unit values. It looks as if indifference to the changes in average prioritarian unit value, then, does not expose prioritarians to the accusation that their position is tantamount to utilitarianism” (Porter 2011, 202). Persson (2011, 308–309) does provide some further argument. He notes that Porter relies on cases in which benefits are either added to or subtracted from the benefits that recipients’ already possess. He concedes that in these cases Porter is right: the prioritarian can maintain indifference to the average moral value of benefits. But he argues that in cases in which there is a radical redistribution of benefits things are otherwise. But it seems to me that the same issue arises in such cases.
Objection. As such, without further argument it seems that the egalitarian must bite a bullet that the prioritarian does not have to bite, namely that leveling down can be an improvement at least in some respect. And if I am right about what I said about Moderate Egalitarianism, the (telic) egalitarian then must allow that it is at least possible that the leveled down state to be better all-things-considered.

**VII. CONCLUSION**

In response to the Leveling Down Objection, defenders of egalitarianism have made several moves. Two have been considered in some detail here. The first is Christiano’s view that claims that a leveled down state is not better even with respect to equality because the value of equality and the value of what is to be equalized are deeply intertwined. The second, made by both Persson and Brown, claims that prioritarianism too is subject to leveling down, and, as such, we should not be so troubled by egalitarianism’s vulnerability to the Leveling Down Objection. I have argued that neither of these responses to the Leveling Down Objection are entirely convincing. This does not mean that (telic) egalitarianism must be abandoned, as one can insist that the Leveling Down Objection is not dispositive: even if (telic) egalitarianism has the consequence that leveled down states can be better in a respect, and therefore possibly better all-things-considered, this is an uncomfortable consequence rather than a decisive objection. The point of this paper, however, is that (telic) egalitarians may have to accept this uncomfortable consequence, as the two attempts discussed here to deny it or lessen its significance are unsuccessful.

---

42 As a last resort, it might be argued that Persson and Brown’s arguments assume telic prioritarianism, leaving it open that deontic prioritarianism is subject to the Leveling Down Objection (or some analogous objection). However, as in the case of egalitarianism, deontic views seem less rather than more likely to be subject to the Leveling Down Objection (or some analogous objection). Further, if telic prioritarianism is not subject to the Leveling Down Objection and telic egalitarianism is, as I have argued, then the egalitarian cannot claim that with respect to the Leveling Down Objection egalitarianism and prioritarianism are equally vulnerable.

43 There are a variety of other approaches worth considering, including the view suggested by Otsuka and Voorhoeve (2009, 182–184) that the strength of a person’s claims to benefits depends on the comparative strength of the claims of others. That, however, must be left for another time.
REFERENCES


**Michael Weber** is Professor of Philosophy at Bowling Green State University, and Department Chair.

Contact e-mail: <meweber@bgsu.edu>
Reason and Political Economy in Hume

ERIK W. MATSON

New York University

Abstract: This paper examines some connections between Hume's epistemology in his Treatise of Human Nature and his political economy. I make three claims: (1) First, I argue that it is the development of Hume's account of the faculty of reason in Book I of the Treatise that leads him to emphasize social science—including political economy—and the humanities over more abstract modes of intellectual inquiry. (2) Second, I argue that Hume's conception of reason has implications for his methodology in political economy. His perception of human reason leads him to deploy a method of qualified generalization that emphasizes the by-and-large nature of theoretical statements. (3) Third, when it comes to policy matters, the method of qualified generalization in theory cashes out in terms of practical maxims. I suggest that two central maxims in Hume's political economy derive from his views of the usefulness of economic liberty and the coordinating nature of the status quo.

Keywords: David Hume, political economy, reason, epistemology, methodology

JEL Classification: B12, B31, B40

I. INTRODUCTION

In 1739, David Hume published Books I and II of his famous Treatise of Human Nature. A year later he published Book III. The Treatise was ill-received, falling, as Hume put it, “dead-born from the press” (1987,

1 Abbreviations: References to the Treatise of Human Nature are to Hume (2007), abbreviated as ‘T’, followed by book, part, section, and paragraph. References to An Enquiry Concerning Human Understanding are to Hume (2000), abbreviated as ‘EHU’, followed by section, part [when one exists], and paragraph. References to An Enquiry Concerning the Principles of Morals are to Hume (1998), followed by section, part [when one exists], and paragraph. References to the Essays, Moral, Political, and Literary are to Hume (1994), abbreviated as ‘EMPL’, followed by page. References to The History of England are to Hume (1983), followed by volume and page. References to An Essay Concerning Human Understanding are to Locke (1975), followed by book, chapter, and section.

AUTHOR’S NOTE: Thanks to Erwin Dekker, Daniel Klein, Thomas Merrill and three anonymous referees for their constructive comments and helpful suggestions.
xxxiv; italics original). After the disappointing reception of the *Treatise* Hume moved on to other projects. He became particularly well-known for his contributions to the social sciences and humanities, becoming a recognized authority in matters of morals, politics, political economy, and history. He was perhaps most famous in his day for his celebrated six-volume *History of England*.

I do not think, as some have argued, that Hume’s post-*Treatise* career and shifting focus towards social science marked an end to his philosophy. Rather, I consider Hume’s post-*Treatise* developments to be consistent with, and even manifestations of, the philosophical explorations in his original volume. In the present essay, I seek to elaborate some connections between Hume’s early thinking in epistemology and his later developments in political economy. Understanding these connections illustrates Hume’s well-developed social outlook and sheds new light on his philosophical approach to economics.

I break down my analysis into three claims. (1) First, I argue that it is the development of Hume’s account of the faculty of reason in Book I of the *Treatise* that leads him to emphasize social science—particularly politics and political economy—and the humanities over more abstract modes of intellectual inquiry. (2) Second, I argue that Hume’s conception of reason has implications for his methodology in political economy. His perception of human reason leads him to deploy a method of qualified generalization that emphasizes the by-and-large nature of theoretical

---

2 In their introductory remarks to Volume 3 of the 1889 edition of *The Philosophical Works of David Hume*, Grose and Green (1889) comment on Hume’s movement from epistemology to social science, speaking of “the suddenness with which [Hume’s] labours in philosophy came to an end” (75). Such thinking, although not uncommon among some of Hume’s earlier interpreters, overstates the independence of Hume’s *Treatise* and his subsequent works and hinges on an overly narrow conception of Hume’s ‘philosophy’ (see Miller 1987, xviii).

3 James Harris (2015) surveys the historical development of views on the relationship between Hume’s *Treatise* and his subsequent writings. Harris argues that Hume should not be read as a systematic thinker but as a man of letters with various philosophical interests and projects: “In later life, Hume nowhere described himself as thinking of his works as a unity of any kind at all. Once he had given up on the *Treatise*, Hume never once presented himself as a systematic thinker, as someone who conceived of his writings in terms of foundation and superstructure, or of core and periphery, or of trunk and branches” (13). I agree with Harris that Hume’s corpus should not be read as a completely unified systematic effort. But this does not mean that there is not significant philosophical, methodological, and conceptual continuity between Hume’s various projects, or that his post-*Treatise* works cannot be read as applications of the principles developed in the *Treatise*. For other similar views see, e.g. Livingston (1984); Miller (1987); Danford (1990); and Merrill (2015). For a slightly different but compatible account, see Immerwahr (1991).
statements. Although we can reason with a high degree of certainty at a
genral and abstract level, the applicability and relevance of such
reasoning to particular matters of fact is much less certain. (3) Third,
when it comes to policy matters, the method of qualified generalization
in theory cashes out in terms of practical maxims. I suggest that two
central maxims in Hume's political economy derive from his views of
the usefulness of economic liberty and the coordinating nature of the
status quo.

II. HUME'S ACCOUNT OF REASON
Hume's account of reason is central to his view of the potentialities of
human understanding. It is also central to his conception of what
philosophers should do—a conception that comes to emphasize
reflection in light of human action rather than the seeking independent
foundations of knowledge. The account develops dialectically
throughout Book I of the Treatise by relating two different concepts of
reason. Hume begins by formulating and employing a narrow concept of
reason largely taken from Locke. He finds that this narrow concept
cannot explain the mechanisms of probable reasoning or our ideas
about causation. The narrow concept of reason cannot justify or explain
our belief in what Don Garrett (1997, 86) calls the “Uniformity Thesis”,
i.e. the proposition that the future will resemble the past (see T 1.3.6.4).
Not wanting to call into question our reliance on experience or the
practice of probable reasoning, Hume moves to reconfigure and broaden
the faculty of reason. His broader version of reason, which includes
probable reasoning, depends upon principles of the imagination by
which the mind transitions from past ideas to present impressions. But
from the perspective of the narrow version of reason Hume recognizes
that it is lacking in epistemological rigor.

The dialectic between the narrow and wide conceptions of reason,
and the resulting question of belief and justification, lies close to the
heart of Hume's philosophy. The narrow and broad concepts of reason
might be said to constitute two “contrarieties of thought which
structure [Hume's] drama of inquiry” (Livingston 1984, 35). This drama
influences the kind of subjects that become central to Hume's
philosophy. Hume's account of reason and its shortcomings leads him

4 For a more developed account of my interpretation of Hume's account of reason and
its relation to the secondary literature see Matson (2018). For broadly similar accounts
see Winters (1979); Baier (1991); Owen (1999).
to shift his philosophy away from metaphysics and towards more practical matters of social science and humanities.

To begin unpacking this account in Hume’s text, it is important to understand that he employs the word ‘reason’ in two major senses. Those senses correspond to: (1) the narrow faculty of reason and (2) the broader faculty of reason. I will refer to the narrow faculty of reason as ‘reason’. Reason1 is a narrow inferential faculty that operates demonstratively by constructing chains of intuitive ideas. I will refer to the wider faculty of reason as ‘reason2’. Reason2 is the broader practical faculty of inference that envelops reason1 and synthesizes it with the principles of the imagination requisite to explain the mechanisms of probable reasoning (see Owen 1999, 203–204).

Early in Book I of the Treatise Hume almost exclusively confines his use of the word ‘reason’ to mean reason1. Apart from some passages in the introduction of the work, Hume seems to mean reason1 almost every time he refers to ‘reason’ (qua faculty) until around T 1.3.11.1. Reason1 corresponds to Locke’s concept of demonstrative reasoning. Demonstrative reasoning in Locke begins with intuitive ideas. Intuition is the mind’s “native Faculty to perceive the Coherence, or Incoherence of its Ideas” (ECHU 4.17.2; italics original). Demonstrative reasoning builds on intuitive ideas by intuiting connections between them and other intuitive ideas. A line of demonstrative reasoning thus forms a chain of intuitively-connected intuitive ideas. By Locke’s definition, demonstrations show “the Agreement, or Disagreement of two Ideas, by the intervention of one or more Proofs, which have a constant, immutable [i.e. intuitive], and visible connection with one another” (ECHU 4.15.1; italics original).

In illustrating the ideas of intuition and demonstration, Locke gives the example of a triangle (ECHU 4.15.1). We can clearly perceive or intuit the idea of a triangle: a closed figure made up of three straight lines. We can intuit the idea of a right angle as the angle made by perpendicular straight lights. Given the idea of a triangle and a right angle, we can demonstrate—perceive the connection between a number of intuitive, intermediate ideas—that the sum of the angles in a triangle equals the sum of two right angles. Such a demonstration is certain by its

5 It should also be noted that sometimes Hume casually uses “reason” in reference to the general activity of reasoning, i.e. the activity of deliberation and inference associated with “having reasons” for doing things.

6 For an elaboration of the Lockean character of reason1, see Owen (1999, especially Chapters 3 and 4).
perception and cannot be subject to question unless one incorrectly perceives the ideas of ‘triangle’, ‘right angle’, ‘straight line’, ‘closed figure’, etc. Demonstrative reasoning thus leads to conclusions that are certain by the law of non-contradiction.

Hume all but explicitly formulates reason1 in a Lockean demonstrative sense within the first passages on the faculty of reason in the Treatise:

A demonstration, if just, admits of no opposite difficulty; and if not just, 'tis a mere sophism, and consequently can never be of difficulty [...] To talk therefore of objections and replies, and balancing of arguments in such a question as this, is to confess, either that human reason [reason1] is nothing but a play of words, or that the person himself, who talks so, has not a capacity equal to such subjects (T 1.2.2.6).

The limitations of reason1 become apparent in the Treatise when Hume turns to consider probable reasoning in T 1.3.6. For Locke, probable reasoning is one of the two modes of the faculty of reason (ECHU 4.17.2). Locke's probable reasoning works on matters of experience by probabilistically inferring from past to present experience—i.e., if $x$ occurred like that in the past, $x$ will, under similar conditions, probably happen like that in the future. Hume similarly conceives of the idea of probable reasoning; but he does not see it to be a matter of reason1.

His own analysis of probable reasoning has both a negative and a positive component. The negative component is simply that probable reasoning is unexplainable by way of reason1: “[P]robable reasoning is not explained by the faculty of reason, traditionally conceived” (Owen 1999, 137). The positive component is Hume’s explication of probable reasoning in terms of his new theory of belief. He conceives of probable reasoning as an imagination-dependent activity facilitated by the mental association of ideas through perceived relations of cause and effect, contiguity, and resemblance (T 1.3.9.3).

For probable reasoning to be explainable by reason1, its requisite presupposition (the Uniformity Thesis) would have to be provable by demonstration. Hume is clear that this is not the case:

[If reason [reason1] determin’d us [in making the inference from a present impression to the idea of an unobserved object], it wou’d proceed upon that principle, that instances, of which we have had no
experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same” (T 1.3.6.4; italics original).

The Uniformity Thesis cannot be demonstrated in that we can “at least conceive a change in the course of nature” (T 1.3.6.5). As Hume says at T 1.2.2.6, a just demonstration does not entail, and cannot possibly entail, objections or counter arguments. It is impossible, for instance, to dispute the demonstration that a triangle has one-hundred and eighty degrees unless one misapprehends one of the ideas—e.g., ‘line’ or ‘degree’—constitutive of the demonstration. If one can conceive an idea contrary to x, then x is non-demonstrable and not reason1-reasonable by definition. The very fact that the mind can entertain the idea of, e.g., the sun not rising tomorrow means that the uniformity of experience is not demonstratively certain (see EHU 4.1.2). Hume concludes, “there can be no demonstrative argument to prove, that those instances, of which we have had no experience, resemble those, of which we have had experience” (T 1.3.6.4; italics original). As every line of probable reasoning logically rests upon the Uniformity Thesis, no line of probable reasoning is provable by demonstration.

Hume recognizes that no one can effectively deny the soundness of probable reasoning in practice. Although the negative component of his argument does raise a “specter of skepticism” that comes to bear in his later reflections (Garrett 1997, 95), he has no interest in dissuading anyone from using probable reasoning, contrary to the interpretations of some of his early critics (Kemp Smith 2005, 22). His point, rather, is that probable reasoning is enabled by a belief-generating operation of the imagination by which ideas are conceived of in “lively” manner “related to or associated with a present impression” (T 1.3.7.5). Put differently, probable reasoning operates within a framework of natural belief or custom that is sustained by the psychological constitution of the imagination. This positive component of Hume’s argument comes to head when he claims, “all probable reasoning is nothing but a species of

---

7 A natural response to this aspect of Hume’s analysis might be to invoke probability and say, “the future may not resemble the past with demonstrative certainty, but it surely resembles the past with probability”. Yet Hume points out that to invoke probability as a resolution to this problem is assuming that which needs resolving: “[A]ll our experimental conclusions proceed upon the supposition, that the future will be conformable to the past. To endeavour, therefore, the proof of this last supposition by probable arguments, or arguments regarding existence, must be evidently going in a circle, and taking for granted, which is the very point in question” (EHU 4.1.19).
sensation. 'Tis not solely in poetry and music, we must follow our taste and sentiment, but likewise in philosophy” (T 1.3.8.12).

Hume recognizes that within the arena of common life, the grounds of probable reasoning are, in a sense, “entirely free from uncertainty” (T 1.3.11.2; italics original). As such, he broadens his working conception of ‘reason’ from reason1 to reason2. He signals this semantic transition towards the end of T 1.3:

Those philosophers [including Locke], who have divided human reason into knowledge and probability, and have defin’d the first to be that evidence, which arises from the comparison of ideas, are oblig’d to comprehend all our arguments from causes or effects under the general term of probability. But tho’ every one be free to use his term in what sense he pleases; and accordingly in the precedent part of this discourse, I have follow’d this method of expression; ‘tis however certain, that in common discourse we readily affirm, that many arguments from causation exceed probability, and may be receiv’d as a superior kind of evidence (T 1.3.11.2; italics original, boldface added).

The emboldened phrase suggests that perhaps Hume has been using ‘reason' in the preceding pages of the Treatise ‘in what sense he pleases’. He has been using ‘reason’ in a sense that differs from Locke and ‘those philosophers’. That sense is reason1, which differs from Locke’s reason (a complex of both demonstrative and probable reasoning) because it excludes probable reasoning (see Millican 1998, 145). But he continues that common sense forcefully tells us that probable reasoning is a superior sort of evidence to other “whimsies and prejudices” of the imagination that probable reasoning at least partially resembles (see T 1.3.9.19 n22). Hume recognizes that the natural-belief framework of probable reasoning has a high degree of practical merit that takes on a type of certainty within the context of ordinary and scientific life (see De Pierris 2001, 377–380). As such, he considers it to be properly designated as ‘reason’ (i.e. reason2).8

8 In a letter to John Stewart, who attacked Hume for his views on the idea of causation, Hume speaks of different sorts of certainty associated with reason1 and reason2. He says that the certainty that results from reason2 deliberations is more sure than other kinds of certainty, but less sure than the demonstrative kind: “[A]llow me to tell you, that I never asserted so absurd a Proposition as that any thing might arise without a Cause: I only maintain’d, that our Certainty of the Falsehood of that Proposition proceeded neither from intuition nor Demonstration; but from another source. That Caesar existed, that there is such an Island as Sicily; for these Propositions, I affirm, we have no demonstrative nor intuitive Proof. Would you infer that I deny their Truth, or
As a conceptual matter, reason2 simply equals reason1 plus probable reasoning and its facilitating principles of the imagination. As probable reasoning is “a species of sensation” (T 1.3.8.12), reason2 is at its core much more a matter of instinct than of strict cognition. Reason2 is “more properly an act of the sensitive, than of the cogitative part of our natures” (T 1.4.1.8; italics original). When Hume gives a definition of ‘reason’ at the end of Book I, Part 3—which is the first explicit definition of ‘reason’ in the Treatise—he defines reason2 thus: “To consider the matter aright, reason is nothing but a wonderful and unintelligible instinct in our souls, which carries us along a certain train of ideas, and endows them with particular qualities, according to their particular situations and relations” (T 1.3.16.9).9

III. HUME’S PATHWAY TO HUMAN THINGS

Hume’s account of reason2 leads him to reorient his philosophy in a way that emphasizes the social sciences and the humanities. This reorientation is a consequence of his conditional acceptance of reason2 as an appropriate concept of ‘reason’. Put differently, it turns out that embedded within Hume’s formulation of reason2 is a prioritization of human things that puts him on the path to investigate, among other things, political economy.10 This comes forth directly in Hume’s dramatic and well-known conclusion to Book 1 of the Treatise (T 1.4.7; hereafter, ‘Conclusion’).

Despite the psychological and practical robustness of reason2, it remains unverifiable from the perspective of reason1. Reason2 depends on principles of the imagination that resemble other “whimsies and prejudices” of thought, e.g. superstition (see T 1.3.9.19 n22); it operates within a practical framework of custom and natural belief. As such, the

---

9 I do not include an interpretation here of Hume’s important section, “Of scepticism with regard to reason”, which immediately follows Hume’s definition of reason2 in T 1.3.16. My basic reading of that section is that after committing to reason2 in T 1.3, Hume unearths some internal problems with reason2. From the simple observation of fallibility, Hume is led to conclude that demonstrative knowledge turns to probability, and successive probability estimations “continually [diminish] the original evidence” and “utterly subvert all belief and opinion” (T 1.4.1.8; for analysis, see Meeker 2000). This problem figures into the drama of T 1.4.7 and the difficulties associated with his final commitment to reason2 and philosophy over alternative ways of knowing (see Matson 2018).

10 For a more extensive interpretation of Hume’s path to human things see Merrill (2015, Chapter 2).
credibility of its applications is not immediately clear. In the Conclusion, Hume reiterates this point, noting that reason2 hinges on “an illusion of the imagination” (T 1.4.7.6). To corroborate reason2, then, he sees that some standard needs to be developed for discriminating between acceptable and unacceptable illusions of the imagination. If we accept reason2, which hinges on the imagination, on what basis can we discriminate against other phenomena produced by the imagination? We cannot entirely reject the imagination as a source of knowledge—if we do, we cannot rely, among other things, on our experience. Rejection of the imagination would leave “not the lowest degree of evidence in any proposition” (T 1.4.7.7). But an open and undiscriminating embrace of the imagination might prove dangerous, leading to wrongheaded thought: “[N]othing is more dangerous to reason than the flights of the imagination, and nothing has been the occasion of more mistakes among philosophers” (T 1.4.7.6).

Hume’s dilemma famously leads to a moment of crisis where he exclaims, “I am confounded with all these questions, and being to fancy myself in the most deplorable condition imaginable, inviron’d with the deepest darkness” (T 1.4.7.8). But this despair is not stable. Hume comes out of his crisis through social engagement: “I dine, I play a game of back-gammon, I converse, and am merry with my friends” (T 1.4.7.9). Upon subsequent considerations, although the memory of his melancholy inclines him to abandon philosophy and “throw all [his] books and papers into the fire”, his curiosity leads him once more to “seclude [himself] from the commerce and society of men” and engage in philosophical matters (T 1.4.7.9). Despite his reasons for skepticism and his understanding of the problems with reason2, he gravitates back towards philosophy.

He emerges from his personal reflections with a newfound commitment to reason2, albeit a conditional one: “[W]here reason [reason2] is lively, and mixes itself with some propensity, it ought to be assented to. Where it does not, it never can have any title to operate upon us” (T 1.4.7.11). His commitment derives both from a sense of public and private usefulness (T 1.4.7.10) and general agreeableness (T 1.4.7.12). This amounts to what Michael Ridge (2003) has referred to as a moralizing of epistemology (see also Owen 1999, 222).

Although the

---

11 When Hume later develops his moral philosophy, he argues that moral approval derives from a spectator’s view of the complex of a virtue or action’s usefulness and agreeableness to the agent and the public.
commitment to reason2 is not based upon a rigorous metaphysics—“we can never have any steady principles, nor any sentiments, which will suit with common practice and experience” (T 1.4.7.14)—Hume views it as conducive to developing “a system or set of opinions, which if not true (for that, perhaps, is too much to be hoped for) might at least be satisfactory to the human mind” (T 1.4.7.14). Despite its problems, reason2 is the basis of the most useful interpretive framework.

The nature of Hume’s commitment to reason2 leads him to prioritize practical matters of social science and humanities. The fact that he accepts reason2 on the basis of its apparent usefulness and agreeableness appears to lead him to shift the practical focus of his philosophy to matters of personal and common human interest:

I cannot forbear having a curiosity to be acquainted with the principles of moral good and evil, the nature and foundation of government, and the cause of those several passions and inclinations, which actuate and govern me. I am uneasy to think I approve of one object, and disapprove of another; call one thing beautiful, and another deform’d; decide concerning truth and falsehood, reason and folly, without knowing upon what principles I proceed (T 1.4.7.12).

Hume discovers through his own journey that human beings are actuated in their reasoning by fundamentally inexplicable desires and principles in their nature. Instead of working to fully rationalize such principles, which he recognizes will inevitably lead to skepticism, he turns to study the arenas in which they are most directly on display.12

It should be noted here that Hume’s commitment to reason2 is not a blanket endorsement of common sense. His recognition of the imaginary constitution of reason2 and the practical nature of his own commitment affects not only the subjects to which he feels reason2 can be justly deployed—i.e. morals, politics, the passions, aesthetics, etc.—but also the manner of just reasoning about such subjects. He later underlines this in his Enquiry Concerning Human Understanding: “[I]n

12 Merrill (2015) nicely captures the essence of Hume’s turn here: “Moral and political topics have pride of place in the inquiries on which Hume now embarks (T 1.4.7.12). This is not because Hume regards those opinions as unshakeable or sacrosanct. It is rather because human beings are the moralizing animals. They spontaneously praise and blame, admire and detest, prefer and disdain. Even philosophers who hold their fellows in contempt cannot avoid praising and blaming. Since the self cannot be studied directly as though it were just another object of science, we must look to what human nature does” (59).
general, there is a degree of doubt, and caution, and modesty, which, in all kinds of scrutiny and decision, ought for ever to accompany a just reasoner" (EHU 12.3.1). The nature of reason2 cautions against a dogmatical spirit, even in the realm of practical affairs. Reason2 operates within a framework of custom and upon a sense of practicality; recognizing it as such should lead to a degree of mitigated skepticism in all reasoning (see T 1.4.7.11; EHU 12).

IV. QUALIFIED GENERALIZATION OF POLITICAL ECONOMY
Hume’s account of reason in the Treatise has important implications for his political economy. At the most basic level, it is his conditional commitment to reason2 that leads him to apply his philosophy to moral, political, and economic (a subset of politics for Hume) matters. But beyond guiding the general direction of his inquiries, his conception of reason also speaks to his view of the proper method of reasoning in political economy. This view turns out to have both theoretical and practical consequences. On a theoretical level, Hume emphasizes reasoning in political economy in terms of qualified generalizations. At a sufficiently general and abstract level we can reason with a high degree of certainty. But the generalization and abstraction of such reasoning renders its conclusions merely “hypothetical”, as John Neville Keynes (2017, 205) uses the term in reference to conclusions of his deductive method. The application of generalized reasonings to particular matters of fact introduces degrees of uncertainty. Even conclusions that are certain in the abstract require qualification in application. On a practical level, Hume’s qualified theoretical generalizations translate into qualified practical generalizations, or

---

13 Scholars have emphasized other connections between the Treatise and Hume's political economy. Eugene Rotwein (2009) connects Hume's theory of the passions (Treatise Book II) to his economic psychology. Carl Wennerlind (2001) links Hume's analysis of obligation and the utility of promises (Treatise Book III) to his monetary theory. Till Grüne-Yanoff and Edward McClenenn (2008) develop a sociological theory of preference change and economic development by drawing on Hume's theory of the passions (Book II). Wennerlind (2011) suggests that Hume's theory of justice as property (Book III) can be adequately understood only in conjunction with his writing on commerce. Margaret Schabas and Wennerlind (2011) draw on the Treatise to argue that Hume accorded economics a higher epistemic status than Newtonian physics—whereas we have no way of observing directly the connecting principle of gravity, Hume thinks we can, by inward reflection, observe causal connections between passions and behaviors.

14 I hereafter mean ‘reason’ as reason2 unless otherwise specified.
practical maxims. These maxims function as presumptions around which particular policy discussions can be organized.

The connection of Hume’s method of qualified generalization to his account of reason, and to the philosophical framework of the Treatise more generally, appears directly in the beginning of the first essay in his Political Discourses, “Of Commerce”:15

The greater part of mankind may be divided into two classes; that of shallow thinkers, who fall short of the truth; and that of abstruse thinkers, who go beyond it. The latter class are by far the most rare: and I may add, the most valuable. They suggest hints, at least, and start with difficulties, which they want, perhaps, skill to pursue; but which may produce fine discoveries, when handled by men who have a more just way of thinking (EMPL, 253; italics original).16

This passage directly parallels a passage from Book I of the Treatise (T 1.4.3.9) and the discussion in the first essay of the Enquiry Concerning Human Understanding (EHU 1). All three passages relate to the philosophical drama in T 1.4.7 discussed in the previous section. The underlying theme is that proper thinking, or what Hume sometimes calls ‘true philosophy’ requires both (1) a recognition of the shortcomings of common or shallow thinking—i.e. rejecting an unreflective acceptance of the standards of custom and common life—and (2) a recognition that the thinking employed by the philosopher to disassemble common thinking itself relies on a frame of natural belief and custom. While reason points to many problems with common beliefs and opinions, it is itself a product of beliefs that cannot be independently verified—reason2 is useful and agreeable but epistemologically lacking from the vantage point of reason1. The just reasoner must take both points into account and walk a line between shallow thought, skeptical paralysis, and dogmatism.

15 Hume’s Political Discourses are a subset of his Essays, Moral, Political, and Literary. The Political Discourses were first published in 1752. The Political Discourses comprised the most successful volume of Hume’s Essays, with a second and third edition published in 1752 and 1754 (see Miller 1987, xi-xviii). The essays within the 1754 edition of Political Discourses are: (1) “Of Commerce”; (2) “Of Refinement in the Arts”; (3) “Of Money”; (4) “Of Interest”; (5) “Of the Balance of Trade”; (6) “Of the Balance of Power”; (7) “Of Taxes”; (8) “Of Public Credit”; (9) “Of some Remarkable Customs”; (10) “Of the Populousness of Ancient Nations”; (11) “Of the Protestant Succession”; and (12) “Idea of a Perfect Commonwealth”.

16 This passage has been referred to by E.C. Mossner (2001) as elucidating Hume’s “broad philosophical approach to the problems of economics” (269).
The first implication of Hume’s notion of true philosophy for political economy is a matter of posture and epistemic humility. The fact that reason is itself to a large degree a product of custom should limit the enthusiasm of the political economist with respect to political progress and policy reform. This idea informs what Frederick Whelan (1985) calls Hume’s “practical conservatism [...] derived from caution prescribed by [his] methodological guidelines for right reasoning” (314). Hume’s practical conservatism interrelates with his practical maxim regarding the coordinating tendencies of status quo arrangements (at least his status quo political arrangements), which I return to below. But the more substantive implications of Hume’s conception of true philosophy pertains to the level at which political economic theory should take place. Hume makes two specific claims about this in “Of Commerce”: (1) First, there should be a strong presumption of falsehood when theory is used to generate specific predictions about particular matters of fact. (2) Second, just theorizing should operate at a high level and offer general but qualified conclusions.

As to the first claim, Hume says, “there are some cases, I own, where an extraordinary refinement affords a strong presumption of falsehood, and where no reasoning is to be trusted but what is natural and easy” (EMPL 254; see also EMPL 47). The logic of this claim rests in part on Hume’s understanding of the philosophical problem with probable reasoning, i.e. the problem of induction, and the psychology of reason. The assumption that the future will resemble the past (the Uniformity Thesis) is not demonstrable yet must be practically committed to in order to reason about matters of fact. But such an understanding undermines the epistemological credibility of specific predictions about matters of fact, a fortiori—since the Uniformity Thesis itself is not demonstrable and all predictions regarding matters of fact rest on the Uniformity Thesis, no predictions are demonstrable. There should, therefore, be a general wariness about reason’s ability to predict particular outcomes. Such wariness strengthens when one recognizes that even once the Uniformity Thesis is granted and internalized into reason, it is unlikely on Hume’s account that an individual mind will properly interpret matters of fact. The mind organizes its experience set

---

17 Ironically, Hume is a bit inconsistent on this matter in practice. He notes in the Treatise that “we are apt not only to forget our scepticism, but even our modesty”, and sometimes appears to do so himself (T 1.4.7.15).
into an interpretive framework on the basis of the perceived relations of contiguity, resemblance, and cause and effect (the settled principles of the imagination that constitute reason). But there is no guarantee that conclusions drawn from this organization will correspond to the truth—every individual’s experience is quite limited. The best to hope for then is the formulation of some general efficient causes that explain the constant conjunction of certain classes of experience. Drawing on this logic, Hume concludes that men, in their particular business affairs, when they form “schemes in politics, trade, [or] oeconomy”, should stay from specific predictions and “too fine” arguments: “Something is sure to disconcert [their] reasoning, and produce an event different from what [they] expected” (EMPL, 254). Reason, by constitution, would seem unfit to credibly offer specific predictions. Unforeseen consequences, perhaps resulting from unanticipated causal forces, abound.

As to the second claim, Hume says:

But when we reason upon general subjects, one may justly affirm, that our speculations can scarcely ever be too fine, provided they be just; and that the difference between a common man and a man of genius is chiefly seen in the shallowness or depth of the principles upon which they proceed. General reasonings seem intricate, merely because they are general; nor is it easy for the bulk of mankind to distinguish, in a great number of particulars, that common circumstance in which they all agree, or to extract it, pure and unmixed, from the other superfluous circumstances (EMPL, 254; italics original).

As previously mentioned, Hume’s arguments here prefigure John Neville Keynes’ (2017, Chapter 7) analysis of the “hypothetical” character of conclusions of the deductive method. On Keynes’ account, we can justly deduce theoretical economic conclusions from general premises, although the relation of such deductions to particular matters of fact depends on the extent to which assumptions of ceteris paribus hold in practice. Keynes (2017), following J. S. Mill, argues, “all laws of causation may be said to be hypothetical, in so far as they merely assert that given causes will in the absence of counteracting causes produce certain effects” (205; italics original). Hume appears to argue along the same lines in “Of Commerce”. The “common circumstances” that he speaks of are causes, which might fail to produce uniform observable effects in practice if they are counterbalanced by “superfluous circumstances” (EMPL, 254). Hume continues to speak of “universal
propositions, which comprehend under them an infinite number of individuals, and include a whole science in a single theorem” (EMPL, 254). The nature of the applicability of such propositions changes due to the varying of local circumstances and the potential presence of counterbalancing causal forces. But like Keynes, Hume does not conclude from this that general and universal propositions have no bearing on particular matters of fact. Rather, he says that the political economist, at least in theoretical matters, should dwell at a level of sufficiently general analysis precisely because such propositions do come to bear, by and large, on the course of events: “[G]eneral principles, if just and sound, must always prevail in the general course of things, though they may fail in particular cases” (EMPL, 254). Analysis on the basis of general principles comports with Hume’s conception of proper thinking so long as the application of those principles is qualified.

The kind of general principles that Hume here has in mind appear to be related to the passions (Rotwein 2009; Schabas and Wennerlind 2011) and the related logic of institutional arrangement. Some of Hume's remarks in “Of Commerce” bear a significant resemblance to his analysis in his famous essay, “That Politics May Be Reduced to a Science”, where he argues, “so great is the force of laws, and of particular forms of government, and so little dependence have they on the humours and tempers of men, that consequences almost as general and certain may sometimes be deduced from them, as any which the mathematical sciences afford us” (EMPL, 15; italics added). Note his emphasis here on the general nature of these deduced institutional consequences. His conclusions in political theory dwells at a general level about the by-and-large effects of particular institutional arrangements, which he formulates in terms of political truths. His theoretical method in political economy follows along similar lines. He appears to think that we can have a high degree of certainty regarding general conclusions about the nature of institutional arrangements, derived from an introspective theory of the passions and the principles of human action.

V. Qualified Generalization in Practice: Maxims of Liberty and Status Quo
Hume's method of qualified generalization in theoretical matters cashes out in policy in terms of qualified practical generalizations or maxims.
These maxims are informed by theory and applied to policy in a by-and-large way that is understood as conducive to the public good (Hume's understood purpose of politics). They can be seen functioning as general presumptions for policy that place a burden of proof on alternative recommendations. As qualified theoretical generalizations might not hold in particular instances, so practical maxims might not always hold in policy; but they are presumptively true, or at least useful, until shown otherwise.

There are two central maxims implicit in Hume that serve as useful lenses through which his political economy can be viewed: the liberty maxim and the status quo maxim. The liberty maxim is the defeasible presumption that liberty in economic affairs, within a settled framework of political authority, best harmonizes private happiness and public interest. Policy that infringes upon economic freedom and dampens commercial spirit in the name of the public good accordingly bears the burden of proof. The status quo maxim is the defeasible presumption that within a stable political order of relatively free government, departures from the status quo have discoordinating or destabilizing tendencies. This second maxim is relevant for Hume given his concern with maintaining political stability and constitutional order in the wake of the tumult of seventeenth-century English politics. It is also related to his view of the difficulties of anticipating the consequences of political reform and posture of moderation.

These two maxims are not the only practical maxims one can find in Hume. Qualified theoretical generalization can lead to other maxims in political economy. But the liberty and status quo maxims are useful touchstones in that they capture the overall aim and spirit of Hume's project. The general thrust of his theoretical economic analysis leads him to advance the liberty maxim; the general prudential spirit of his politics, combined with his view of the frailties of reason, leads to the status quo maxim.

---

18 I take the general concepts of the liberty and status quo maxims and their operation as defeasible presumptions in policy from Klein (2012).
19 Here is how Hume defines free government: “The government, which, in common appellation, receives the appellation of free, is that which admits of a partition of power among several members, whose united authority is no less, or is commonly greater than that of any monarch; but who, in the usual course of the administration, must act by general and equal laws, that are previously known to all the members and to all the subjects” (EMPL, 40–41).
V.I. The Liberty Maxim

The backbone of the liberty maxim comes from Hume’s spontaneous order sensibilities that are first developed in the Treatise and extend all the way through to his Dialogues Concerning Natural Religion (Hamowy 2005, 46–49). In the Treatise Hume articulates that the various institutions of society, from conventions of property and the rules of justice to language and money, are functions of human nature and cultural evolution—they are not consciously designed or calculated (see Hayek 1963). Hume has some confidence in the robustness of these institutions once they are established, especially property and the rules of justice, in that they are selected for, in an evolutionary sense, on the basis of social usefulness. Property and the rules of justice emerge as individuals within a social group recognize that it is in their own long-term interest to abstain from taking what is conventionally understood as the property of others. The rules of justice rely on self-interest to channel individual’s more violent passions and desires into calmer and peaceable pursuits of personal advantage (see Hirschman 1977, 24–26). Hume concludes from such thinking, “after the agreement for fixing and observing of this rule [the distinction of property] there remains little or nothing to be done towards settling a perfect harmony and concord” (T 3.2.2.12). This is the basic logic of the liberty maxim: within a settled framework of the rules of justice, which are merely the protection of property and enforcement of contract, economic liberty is presumptively most conducive to both private and public good.

Hume builds out the liberty maxim from his early thinking on spontaneous order in the Treatise throughout his Essays. The maxim comes across strongly in the more directly economic essays in the Political Discourses. Indeed, one of the major themes of those essays is the moralizing of commercial spirit upon the view that commercial activity naturally arising from economic liberty leads to human flourishing and happiness (Schabas 2014). Hume leverages theoretical generalizations about the benefits of commerce throughout the essays into generalized policy advice in terms a presumption of liberty.

Immediately after the discussion of qualified theoretical generalization in the introduction to “Of Commerce”, Hume implicitly argues for something close to liberty maxim by asserting: (a) the power of the state depends on the commercial activity of its citizens; (b) the happiness of its citizens depends on the power of the state to protect their commercial activity; and (c) therefore both the power of the state
and happiness of citizens are inextricably linked to flourishing commerce. But he immediately qualifies his theory: “This maxim is true in general; though I cannot forebear thinking, that it may possibly admit of exceptions, and that we often establish it with too little reserve and limitation” (EMPL, 255). It is possible that the interest of the state and the happiness of its citizens might diverge. In some cases, the state might be made more powerful by limiting commercial activity: “[T]here may be some circumstances, where the commerce and riches and luxury of individuals, instead of adding strength to the public, will serve only to thin its armies, and diminish its authority” (EMPL, 255). But Hume’s subsequent analysis of the applicability of his maxim in various historical circumstances serves to place the burden of proof on anti-commercial economic policy.

The theoretical basis for the liberty maxim strengthens in the subsequent essays. Perhaps the most comprehensive case is made in the second essay of the Discourses, “Of Refinement in the Arts”, where Hume develops an optimistic account of what he sees as an “indissoluble chain” between industry, knowledge, and humanity, and good governance (EMPL, 271). Based on his theory of the passions, he sees economic liberty as a centerpiece of individual well-being as it enables and promotes industry and an active life. Industry has positive social effects in that it leads to innovation and refinement in the mechanical arts and sciences. Refinement in the arts and sciences leads to sociability and confers positive externalities on the liberal arts, which jointly lead to progress in the knowledge and practice of virtue. He connects public interest to the chain: “[I]ndustry, knowledge, and humanity, are not advantageous in private life alone: They diffuse their beneficial influence on the public, and render the government as great and flourishing as they make individuals happy and prosperous” (EMPL, 272; italics original). Hume thinks that the whole process will have a positive effect on political spirit and the nature of policy-making, encouraging moderation and the “advantages of humane maxims above rigour and severity” (EMPL, 273).

In “Of Money”, “Of Interest”, “Of the Balance of Trade”, and “Of the Jealousy of Trade”, Hume drives home the theoretical basis of the liberty maxim and its policy implications. Among other things, he shows how commercial advantages, absent trade restrictions, are conferred from country to country. He walks through the logic of what is now referred to as the price-specie flow mechanism, which speaks to the
insignificance of trade balance figures even from a concern with stocks of bullion—outflows of precious metals will lower domestic prices, place upward pressure on exports relative to imports, and attract precious metal inflows. The policy implications of these theories perhaps come forth most directly in “Of the Balance of Trade” where he exclaims against both domestic and international trade restrictions:

From these principles we may learn what judgment we ought to form of those numberless bars, obstructions, and imposts, which all the nations of Europe, and none more than England, have put upon trade [...] Our modern politics [...] adopt a hundred contrivances, which serve no purpose but to check industry, and rob ourselves and our neighbors of the common benefits of art and nature (EMPL, 324).

Interestingly, however, as with his statement of the liberty maxim in “Of Commerce”, Hume almost immediately qualifies his assertion, in keeping with the qualified nature of his theoretical reasoning and his moderation in application. He admits that not all taxes on foreign goods are “prejudicial or useless” (EMPL, 324). He offers moderate support for taxes on German linen because he thinks that such support encourages domestic manufacturers; he supports a brandy tax in that it increases the relative demand for rum and supports British colonies. Later, in his History of England, we see a similar statement of the liberty maxim followed by a qualification. The initial formulation, as in “Of the Balance of Trade”, is quite direct. He says it should be “the constant rule of the magistrate, except, perhaps in the first introduction of any art [...] to leave the profession to itself, and trust its encouragement to those who reap the benefits of it” (H 3.135; see also H 3.78). But in the next paragraph he says that there are, however, some arts and professions that might perhaps call for some kind of regulation or subsidization, particularly when the provision of public goods is concerned.20

V.II. The Status Quo Maxim
The second central practical maxim in Hume is the status quo maxim. The status quo maxim, again, is the defeasible presumption that within a stable political order of free government, departures from the status quo have dis-coordinating tendencies. Unlike the liberty maxim, the status quo maxim is not directly related to the substance of Hume's

20 He speaks specifically of “finances, armies, fleets, and magistracy”, followed by an interesting case for state-sponsored religion (H 3.135).
economic policy recommendations. Rather, it serves as a kind of meta-qualification for the application of any theoretical principles or their derived practical maxims. The qualified generalization within the status quo maxim is that existing policies within an established free government are beneficial, regardless of their individual theoretical or practical merit, because they contribute to the established political order. Reforming the status quo bears a burden of proof in that no individual practical improvements in economic policy are worth risking the stability of the polity.

The theoretical basis for this maxim is twofold. First, and most importantly, without a stable framework of political authority, the benefits of economic liberty will not be realized. Second, the mean by which reforms are ascertained—the faculty of reason—is imperfect, prone to misinterpretation and a misapprehension of consequences. This second point relates to the general attitude of epistemic humility implied by Hume’s concept of just thinking.

As to the first point, economic policy on Hume’s account needs to be considered in the context of its contribution to established political order. The conceptual logic of the liberty maxim and the indissoluble chain of industry, knowledge, humanity, and good governance falls apart without a basic frame of authority. Economic reforms, even liberalizing ones, might have significant disconcerting effects that are potentially harmful to their own cause. The status quo of authority can challenge even the liberty maxim in that individual cases of economic liberalization might weaken the framework of political authority, which in fact threatens liberty itself. Hume directly articulates this point: “[L]iberty is the perfection of civil society; but still authority must be acknowledged essential to its very existence: and in those contests, which so often take place between the one and the other, the latter may, on that account, challenge the preference” (EMPL, 40). That which is inimical to stable, generally liberal, political authority is inimical to liberty in the general course of things. Much like the rules of justice, where we see the usefulness of the whole scheme despite single acts of enforcement that might offend our moral sentiments, we might tolerate established practices that impinge upon direct liberty in that they are somehow integral to the established political order, which is good for liberty on the whole.

Hume’s sensibilities on these matters derive in large part from his historical context. In the wake of the political and social turmoil in
England in the seventeenth century, there was genuine concern among intellectuals that the established political framework upholding and enabling commercial refinement and social progress would decay and lead to social regress (Schabas 2014, 981–983). The concern was not without reason. As late as 1745, there were violent Jacobite uprisings in Scotland attempting a Stuart restoration. In light of such events, Hume worked to develop an “establishment political philosophy […] to give the established regime, the Revolution Settlement, the Hanoverian succession, the respectable intellectual foundation which, in the ‘fashionable system’, it had not got” (Forbes 1975, 91). This so-called establishment political philosophy comes across clearly in the Discourses, especially in the essay “Of the Protestant Succession”, which is explicitly concerned with philosophical justifications of the political legitimacy of the Hanovers in light of the revolution of 1688.

There are a number of examples throughout Hume’s work of him expressing some sympathies for theoretically poor status quo economic and political arrangements on account of their contribution to political stability. In the History of England, in his treatment of the parliamentary remonstrance against the prerogative of Charles I, he notes a list of political and economic abuses, including: “The forced loans: The illegal confinement of men for not obeying illegal commands: The violent dissolution of four parliaments: The arbitrary government which always succeeded: The questioning, the fining, and imprisoning of members for their conduct in the house: The levying of taxes without consent of the commons” (H 5.351). But despite his understanding of the harmfulness of these policies, and the fact that they resulted from crown prerogative, he sympathizes with the royalist position that prerogative should not be fully restricted. The reason for his sympathies is that he understands that restricting authority might, in some cases, lead to populist enthusiasm and animus against political authority, which could, in certain circumstances, have bad consequences for liberty (which it did, at least in the short run, with the execution of Charles I and the rise of Cromwell). From these reflections he draws a general conclusion:

To prevent such an evil, no expedient is more proper, than to contain ourselves within the bounds of moderation, and to consider, that all extremes, naturally and infallibly, beget each other. In the same manner, past usurpations of the crown, however excusable on account of the necessity or provocations whence they arose, have excited an immeasurable appetite for liberty; let us beware, lest our
encroachments, by introducing anarchy, make the people seek shelter under the peaceable and despotic rule of a monarch (H 5.356).

A policy that seems theoretically sound, for instance a restriction of arbitrary economic power of the sovereign, might go awry in practice and subvert its own aims. Restricting a degree of arbitrariness (in some contexts) might excite enthusiasm against political authority in general, which could degenerate into chaos and crisis, subverting the framework in which economic liberty and commercial activity subsists.

The second basis for the status quo maxim pertains directly to Hume’s account of reason. His reflective commitment to reason despite its shortcomings in the Treatise motivates epistemic humility and moderation. Simply put, we should be hesitant and reserved about the potentialities and capabilities of reason given a basic understanding of its fallibility. As Hume argues, “we must, therefore, in every reasoning form a new judgment, as a check or controul on our first judgement or belief” (T 1.4.1.1). Such checking is particularly important in political matters given their complexity—“political questions are infinitely complicated” (EMPL, 507)—and importance for coordinating social affairs. Hume sees political orders, at least the English political order, as emerging haphazardly, as a consequence of the unintended and unforeseen effects of various historical accidents. The essential and contributory elements of the political order are therefore difficult to disentangle, rendering the art of predicting effects of particular policy reforms on the polity as a whole to be tenuous at best.

V.III. Liberty and Status Quo in Conversation

Before concluding, a few brief general comments on the practical relationship between the liberty maxim and the status quo maxim are in order. Both maxims in Hume derive from qualified theoretical generalizations, the liberty maxim from a theory of spontaneous order and the usefulness of economic liberty, the status quo maxim from a

---

21 Hume articulates a similar logic at H 1.169 where he contrasts the “seeming liberty or rather licentiousness of the Anglo-Saxons” with “true liberty”. These general sensibilities are echoed in his attitude towards the radical populist “Wilkes and Liberty” movement in his own time (Raynor 1980). For an elaboration of Hume’s conception of “true liberty” and its relation to political authority, see Klein and Matson (forthcoming).

22 For a comment on the relation and cases of tension between liberty and status quo, see Klein (2012, 255).
theory of the coordinating nature of established political order. Both maxims are therefore defeasible: they are presumptions that locate the burden of proof but that can be overturned should that burden of proof be met. But what happens if they conflict and place burdens of proof in the opposite direction?

In any given economic policy deliberation, there are two possible relations between liberty and status quo. First, they might be in agreement and jointly place the burden of proof on the proposed reform. This will be the case in instances where a proposed policy would depart from the status quo arrangement and impinge upon liberty. Proposals of this type, e.g. economic interventions, trade obstructions, etc., would bear a significant burden of proof from a Humean perspective in that they simultaneously run the gambit of dampening the benefits of commerce and destabilizing the political order. In other words, both the liberty and status quo maxims tell against them. Second, a policy proposal might be recommended on the basis of the liberty maxim but run counter to the status quo maxim. This will be the case in instances of policy liberalization, e.g., removing trade barriers, industry supports, usury caps, etc. The individual logic of such situations will vary. But we can perhaps conclude that the maxim that prevails, i.e. places the burden of proof, will depend upon the degree and continuity of political order. In a context such as Hume’s, where the established political framework is relatively new, it may well be the case that the status quo maxim overturns the liberty maxim for a time, placing the burden of proof upon liberalization efforts. In such cases, the preservation of the burgeoning political order will be more important, and in the long-run more liberalizing, than any individual liberalization efforts. In more established and stable political orders, however, the liberty maxim will likely overturn the logic of status quo, shifting the burden of proof towards those who would defend illiberal status quo arrangements.

The more general point, however, is that there is no algorithm or hard-and-fast rule for resolving conflicts between liberty and status quo, nor for determining whether or not a burden of proof has been sufficiently met to overturn either of the standalone presumptions. Such deliberations will often be matters of taste, as Hume might put it. This is not to say that they are arbitrary and subjective, but rather that discernment in particular instances will depend on one’s practice, education, and exposure to history. Hume would say that judgment in
VI. CONCLUSION

I have argued in this essay that there are important connections between Hume’s early work in epistemology and his later writings in political economy. Hume’s dialectical account of the faculty of reason in Book I of his Treatise leads him to skepticism, which is resolved by a conditional commitment to reason so long as it is applied to things that appear useful and agreeable from the vantage point of social life. Among such things, morals, politics, and economics loom large. Apart from prompting him to direct reason to human things, Hume’s account of reason turns out to have methodological implications for his political economy, as is made clear in the introduction to his essay, “Of Commerce”. Hume’s views on the imperfections of reason lead him to deploy a theoretical method of qualified generalization in his political economy, which cashes out in matters of policy in terms of practical maxims. Two central practical maxims one can see in Hume’s political economy are the liberty maxim and the status quo maxim, which relate to his view of the benefits of economic liberty and the coordinating tendencies of status quo arrangements insofar as they contribute to political stability.

As an historical matter, this account of Hume portrays him, if not as an entirely systematic thinker, as a thinker with significant methodological continuities across important and influential strands of his thinking. It seems not to be the case that Hume abandoned philosophy after the Treatise, but rather that he envisioned a proper application of philosophy, given his epistemological explorations in that volume, to pertain to social sciences and humanities—things naturally of interest and importance from the vantage of common life. As a more philosophical matter, understanding connections between Hume’s epistemology and his method of qualified generalization opens the way for further case studies into the merits and limits of qualified theoretical reasoning and practical maxims in political economy, both on their own terms and in relation to the methodological contributions of thinkers after Hume.
REFERENCES


**Erik W. Matson** is a Postdoctoral Fellow at New York University, Department of Economics. He studies the history and philosophy of economic thought, focusing on David Hume and Adam Smith. Some of his recent work has been published in History of European Ideas, The Adam Smith Review, and the Journal of Scottish Philosophy. Contact e-mail: <ewmatson@gmail.com>
On the Practical Impossibility of Being Both Well-Informed and Impartial

SVEINUNG S. SIVERTSEN
University of Bergen

Abstract: Adam Smith argued that the ideal moral judge is both well-informed and impartial. As non-ideal moral agents, we tend only to be truly well-informed about those with whom we frequently interact. These are also those with whom we tend to have the closest affective bonds. Hence, those who are well-informed, like our friends, tend to make for partial judges, while those who are impartial, like strangers, tend to make for ill-informed ones. Combining these two traits in one person seems far from straightforward. Still, if becoming well-informed is, as Smith also claims, a matter of imaginative perspective-taking with the “person principally concerned” (TMS, I.i.1.4, 13), it might be possible to become well-informed without the emotional entanglement that comes from any actual proximity to those we judge. Against this intuition, I use Construal Level Theory to show that the tension between being well-informed and impartial is likely to persist even if we take any actual proximity out of the equation. I end by discussing some implications of this, and suggest that we should consider revising the ideal to accommodate them.

Keywords: Adam Smith, impartial, well-informed, conventionalism, Construal Level Theory, proximity, bias, perspective-taking

JEL Classification: A110, B120, B310, B400

I. INTRODUCTION

In The Theory of Moral Sentiments, Adam Smith argues that an ideal moral judge must be both impartial and well-informed (TMS, III.2.32, 150-151).1 This sounds intuitively right, and much of the secondary

1 This and all subsequent references to The Theory of Moral Sentiments, abbreviated as ‘TMS’, will be to the Cambridge edition (Smith 2002). References include, in this order, part, section (if applicable, in lower case roman numerals), chapter, and paragraph (both in Arabic numerals).

AUTHOR’S NOTE: Thanks to Preben Sørheim and the other people at the Department of philosophy at the University of Bergen who have commented on earlier drafts of this article. Thanks also to the two anonymous reviewers for their extremely helpful and constructive criticism, and to the editors at EJPE for their support.
literature on TMS assumes that moral agents can in fact live up to this ideal and be both impartial and well-informed. As an example, take the debate about conventionalism: this debate is concerned with whether an individual can transcend the conventional standard of propriety into which she or he has been socialised, by means of reflecting on what an impartial and well-informed spectator would approve of.2 As such, the debate rests on the assumption that persons can, in fact, adopt the perspective of such a spectator.

Fonna Forman (2010) has challenged this assumption by pointing out a tension between the two traits of this spectator. According to Smith, being impartial requires a certain affective distance, while being well-informed typically presupposes a degree of physical proximity. However, physical proximity tends to breed affective proximity, thus precluding impartiality. Hence, friends may be the best-informed spectators, but they also tend to be the most partial. Conversely, strangers may be the most impartial spectators, but they also tend to be the least-informed. Smith’s ideal, Forman complains, is “essentially Janus-faced”, requiring us to be close and distant at the same time (Forman-Barzilai 2010, 159).

Still, Forman has left the door ajar for those who might want to salvage the impartial spectator—and the reason to want to do so extends beyond the debate about conventionalism. The impartial spectator also represents a non-ideal alternative to later Ideal Observer Theories. To its adherents, the fact that the impartial spectator is a realistic ideal, something we can in fact adopt and inhabit, is also what makes it a usable ideal for everyday moral deliberation. If the impartial spectator turns out not to be realistic in this sense, then it also ceases to be an attractive alternative.

Forman, I hinted, has left the door ajar. In her telling of the tension between understanding and impartiality, the tension is primarily mediated by proximity in physical space. However, according to Smith, the main tool of the moral judge is her imagination, and our imagination can “bring home” (TMS, I.i.4.6, 26) to us the cases of people we have never met. With physical space thus taken out of the equation, might not the tension also be eased or eliminated? Unfortunately, no—or so I will argue. The reason is that the proximity bias in our moral judgements also extends to perceived or psychological proximity.

2 To take but three examples, see Samuel Fleischacker (2011), Carola Freiin von Villiez (2011), and Christel Fricke (2011).
Psychological proximity, moreover, is the inevitable result of the imaginative process involved in becoming well-informed—of “bringing the case home” (TMS, I.i.1.4, 13).

Hence, the tension between understanding and impartiality runs deeper than previously recognised. This presents a serious problem both for those who argue for the convention-transcending potential of the impartial spectator, and for those who see Smith’s ideal as a realistic alternative to Ideal Observer Theories for everyday moral deliberation. I end by briefly considering how we might respond to this problem.

II. PROXIMITY BIAS
That spatial relations shape our moral judgements has been a topic in ethics at least since the ancient Stoics. The Stoics noted how we tend to sympathise more readily with those who are close to us, so that the strength of our moral concern declines progressively as we move outward in what may be visualised as concentric circles representing, simultaneously, physical and affective proximity. First, there is the inner circle of close family and friends, then, moving outwards, acquaintances, neighbours, compatriots, and, outermost, humanity as a whole. These are the “circles of sympathy” (Forman-Barzilai 2010, 8–9). Smith adopts the Stoic image, and the connection between physical proximity and moral concern crops up in Smith’s choice of the “spectator” as the model moral judge (TMS, I.i.1.4, 13). For, in Smith’s standard case of moral judgement, the spectator is physically close enough to see with her own eyes what the “person principally concerned” is going through (TMS, I.i.1.4, 13; Forman-Barzilai 2010, 142).

Yet, for all the importance of the faculty of vision to Smith’s model (Griswold 2010, 68), it is, Smith argues, only with the help of our imagination that what we see can be transformed into an understanding of what someone else is going through—namely through imagining what we ourselves would feel if we were them:

Though our brother is upon the rack, as long as we ourselves are at our ease, our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations. Neither can that faculty help us to this any other way, than by representing to us what would be our own, if we were in his case. By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it
were into his body, and become in some measure the same person
with him, and thence form some idea of his sensations, and even feel
something which, though weaker in degree, is not altogether unlike
them. His agonies, when they are thus brought home to ourselves,
when we have thus adopted and made them our own, begin at last to
affect us, and we then tremble and shudder at the thought of what
he feels. (TMS, I.i.2, 11–12; my emphases)

Our senses prompt our imagination, which in turn carries us 'beyond
our own person', 'bringing home' and bringing alive to us the agony of
'our brother'. Physical proximity, that is, helps determine both how
likely we are to try to imagine ourselves in someone else's shoes, and
how likely we are to succeed in so doing. The former because our senses
continually prompt our sympathetic imagination, and so what is close
enough to be sensed will also tend to lay claim to most of our
sympathetic bandwidth. The latter because direct sensory perception
also affords us with the "minutest incidents" Smith thought vital to well-
formed moral judgement (TMS, I.i.4.6, 26; Forman-Barzilai 2010, 143,
159)—the kind of evocative detail we seldom imagine spontaneously by
ourselves, but which appears to play an important role in determining
whether we are able to feel, as opposed to merely think, what it must be
like to be in someone else's shoes (Gilbert and Wilson 2007, 1345; see
also Levine, Lench, and Safer 2009, 1060).

Therefore, even though physical proximity is neither necessary nor
sufficient for sympathy (Forman-Barzilai 2010, 144–145), we will,
typically, be both more ready and more able to sympathise with those
who are physically close enough to be seen or sensed. As Forman puts
it, “the dominant strain in Smith's thought regarding physical space is
that physical proximity makes contact likelier and sympathy therefore
more precise” (Forman-Barzilai 2010, 145–146). Repeated interactions
tend to reinforce this pattern, forging friends from acquaintances
(Forman-Barzilai 2010, 153; Nahemow and Lawton 1975; Newcomb
1956). And from friends, the best-informed spectators (TMS, I.i.4.9, 28;
Forman-Barzilai 2010, 159; see also Stinson and Ickes 1992).

However, it is not from our friends that we can expect to learn the
real truth of how an impartial spectator would see us. Friends are too
interested in our well-being to judge us with the degree of “indifference”
that is required for impartially (TMS, II.i.2.2, 81; III.2.32, n22, 152); they
care too much about us as individuals to see us as “but one in the
multitude in no respect better than any other in it” (TMS, II.ii.1.1, 97;
III.3.4, 158). Consequently, they tend to be too *indulgent* with our foibles and transgressions (Forman-Barzilai 2010, 159). For truly impartial judgements, we therefore have to turn to strangers, who have no special interests in us (TMS, I.i.4.9, 28; III.3.38, 178; Forman-Barzilai 2010, 159). The trouble is that what the stranger gains in terms of impartiality by her affective distance, she also loses in terms of understanding by the physical remoteness that grants her this indifference (Forman-Barzilai 2010, 159).

Friends, then, tend to be well-informed but partial, while strangers tend to be impartial but ill-informed. It is far from clear how any single person could combine the two into one and be both impartial and well-informed at the same time (Forman-Barzilai 2010, 160).

Far from clear—but perhaps not impossible. For, if it is true that the tension between understanding and impartiality is primarily mediated by physical proximity, what happens if we bypass physical space entirely? Could the tension then also be eliminated? If the impartial spectator familiarises herself with the case of a distant stranger, not by coming into close physical contact with him, but through learning about his situation and vividly imagining it from a distance—would her well-informed judgements of him necessarily end up being partial as well? This seems much less certain. It is one thing to say that close friends typically both understand and are too understanding of each other, quite another to say that knowledge and affection are entangled in such a way as to make the trade-off between understanding and impartiality inevitable.

It may seem less certain, but, as we shall see, it is likely true. The reason lies hidden within the metaphor of bringing the case home. Let us therefore begin by examining that metaphor more closely.

### III. Bringing the Case Home

Throughout TMS, Smith employs the metaphor of bringing the case home as a shorthand for the imaginative work involved in becoming well-informed. In Part I, where he establishes the basics of his moral psychology, the metaphor is embedded in a description that makes clear just how demanding this work can be:

He must adopt the whole case of his companion with all its minutest incidents; and strive to render as perfect as possible, that imaginary
change of situation upon which his sympathy is founded. (TMS, I.I.4.6, 26)³

To understand this quote, we need some background on Smith’s moral psychology. Smith begins TMS by noting that people are not wholly selfish—we do in fact care about each other (I.i.1.1, 11). The psychological principle that underlies this other-care is, according to Smith, our capacity for “sympathy”, which he defines as “fellow-feeling with any passion whatever” (I.i.1.5, 13), or what we today would ordinarily call empathy: by imaginatively taking your perspective, I get some sense of what you must be going through.⁴

However, our capacity for sympathy is not simply a capacity to share other people’s feelings. Rather, it is a capacity that allows us to evaluate those feelings by comparing them to how we ourselves feel when we imagine being in their place:

The man who resents the injuries that have been done to me, and observes that I resent them precisely as he does, necessarily approves of my resentment. The man whose sympathy keeps time to my grief, cannot but admit the reasonableness of my sorrow [...] he who laughs at the same joke, and laughs along with me, cannot well deny the propriety of my laughter. (TMS, I.I.3.1, 20)

More generally, if you observe me reacting to something, you are first disposed to look for that something to explain my behaviour (I.i.1.9, 14–15). When casting about for a candidate cause, you may find one that you can imagine reacting to similarly if you were me. If this also appears to be the actual cause of my reaction, you will find that you share my appraisal of the situation, and hence that you sympathise with my sentiments. If you were me, you would have reacted (more or less) as I have in fact done, and hence you also approve of my sentiments as “suitable to their objects”, as appropriate (I.i.3.1, 20).⁵ If, on the contrary, you are unable to sympathise with my sentiments, finding them

³ See also TMS, I.i.2.12; I.i.1.4, 13; I.i.1.7, 14; I.i.2.6, 20; I.i.3.1, 20; I.i.3.9, 23; I.i.5.2, 29; I.I.2.5, 83; I.I.3.3, 85; I.I.5.3, 88; I.I.5.3a11, 91; I.I.5.6, 89; I.I.2.1, 128; V.2.5, 237; VI.i.3.1, 276; VII.i.1.4, 374.
⁴ ‘Ordinarily’ because the precise meaning of that term is contested in ways which need not concern us here (see for example Coplan 2011; Jahoda 2005).
⁵ A spectator may of course also, upon recognising that she sympathises with the agent, be appalled at her own reaction, worrying, say, that she harbours the same racist attitudes as expressed by the agent (Broadie 2006, 172). In so doing, however, she is taking the perspective of a spectator on her own reaction, repeating the process of sympathetic spectatorship.
“unsuitable to the causes which excite them” (I.i.3.1, 20), you disapprove of my sentiments as inappropriate to the degree that you find you would have reacted differently:

If my animosity goes beyond what the indignation of my friend can correspond to; if my grief exceeds what his most tender compassion can go along with; [...] if I laugh loud and heartily when he only smiles, or, on the contrary, only smile when he laughs loud and heartily; in all these cases, as soon as he comes from considering the object, to observe how I am affected by it, according as there is more or less disproportion between his sentiments and mine, I must incur a greater or lesser degree of his disapprobation: and upon all occasions his own sentiments are the standards and measures by which he judges of mine. (TMS, I.i.3.1, 20–21)

Hence, all our judgements about what is appropriate—that is, moral—behaviour are ultimately predicated on the presence or absence of sympathy. If I sympathise with you, I approve, and I judge your actions appropriate. If I do not sympathise with you, I disapprove, and I judge your actions inappropriate.

If Smith is right about this relationship between sympathy and judgements of propriety, then it follows that we must do some serious imaginative footwork to qualify a moral judge.

We must, that is, always be ready to bring home to ourselves the case of the other, be ready to imagine their situation ‘with all its minutest incidents’ to feel how we would react if we were them. If we do not, we are not in a position to say whether our sympathy or lack thereof really corresponds to the propriety or impropriety of the action in question; our judgement might be wrong because we are ill-informed. As we shall see, it is this bringing home that creates the deepest tension between understanding and impartiality.

IV. CONSTRUAL LEVEL THEORY

The perspective-taking component of what Smith called sympathy, and we today call empathy, is often construed as a kind of going visiting, in the sense that we imaginatively transpose ourselves into the situation of the person whose perspective we take. We talk of “walking” or being in another’s “shoes” (Darwall 1999, 143), seeing with another’s “eyes”

6 There is some empirical evidence to support Smith’s account of the operation of sympathy; see Hepach and Westermann (2013).
In this light, the metaphor of bringing the case home seems a curious companion, since it picks out a movement in the opposite direction. Research on the interrelations between the level of detail in which we consider something and the perceived closeness of that thing sheds some light on why the two apparently opposed movements involved in sympathising—going visiting and bringing home—nonetheless hang together. It all comes down to movements along what Yaacov Trope and Nira Liberman call “dimensions” of “psychological distance” (2010, 440).

The idea of psychological distance originates in a simple observation: we only directly experience the here and now, and yet things that are not here and now—for example, memories of the past and plans for the future occupy a large part of our thinking (Trope and Liberman 2010, 440). Our capacity for such “mental time travel” has been extensively studied (see for example Suddendorf and Corballis 2007, 299), but we also spend a considerable amount of time thinking about things that are spatially removed from us, about other people, and about things that could have happened or are more or less likely to happen. Trope and Liberman therefore suggest conceptualising psychological distances as extending outward from our selves in the present, in four interrelated dimensions: temporal, spatial, social, and hypothetical (Trope and Liberman 2010, 442–443). In a manner of speaking, then, our minds are multidimensional psychological spaces.

This psychological space is populated by what we think about—or, in the parlance of social psychology, by mental construals (Trope and Liberman 2010, 440). To take a non-moral example: I can think about the holiday to Italy that my girlfriend and I have planned for next year, and my trip to Germany coming up in a few days’ time. Thinking about each of them brings them both into my psychological space, but, crucially, I perceive the trip to Italy as temporally more distant than the one to Germany.

A central claim of Construal Level Theory (CLT) is that this difference in perceived temporal distance relates to the way I think about each trip, to the way I construe it. When I think about the future trip to Italy, I will tend mostly to focus on general features such as enjoying the sun, exploring, and so on. My mental construal of the temporally distant trip will tend to be abstract and involve “high level” features of what I imagine (Trope and Liberman 2010, 441). In contrast,
thinking of my trip to Germany two days from now, I will tend to focus on details like what to pack, and at what time I have to leave from home to reach the airport in time. My mental construal of the temporally closer trip will tend to be concrete and involve “low level” features of what I imagine (Trope and Liberman 2010, 441).

However, if I put aside the details of my trip to Germany for the moment and think instead of the purpose or goal of the trip—typical high-level construals—the trip may appear temporally more distant, like there is still plenty of time to prepare everything (an effect known as the ‘planning fallacy’—see Kruger and Evans 2004; see also Nussbaum, Liberman, and Trope 2006, 159). Conversely, if I get my act together and actually sit down to book tickets to Italy, the shift in focus from high-level pleasantness to the low-level details of our itinerary may make next summer seem a little less distant. In other words, the link between construal level and psychological distance is bidirectional (Liberman et al. 2007). Therefore, our perceived or psychological distance to something is the result of a dynamic interaction between factors outside our control, like actual temporal distance, and something over which we have some control, namely, the way we construe the thing we are thinking about.

Thus, going visiting and bringing home both point to a reduction in psychological distance, a traversing of the psychological space between the spectator and the person principally concerned. If we translate Smith’s injunction to “adopt the whole case of [our] companion with all its minutest incidents” (TMS, I.i.4.6, 26) into CLT-ese, it reads as an instruction to traverse the psychological distance between ourselves and our companion by forming a detailed, low-level construal of her situation. However, doing so is not without consequence for our ability to judge her impartially.

V. CONSTRUAL LEVEL THEORY AND MORAL JUDGEMENT
Research has revealed that construal level affects moral judgements (for a recent, if incomplete review, see Mårtensson 2017). The first study to
examine this connection, by Eyal, Liberman, and Trope (2008), found that people would judge moral transgressions more severely if the transgressions were construed as taking place in the more distant future (see their Study 2).\(^9\) This may surprise you: should we not react more strongly towards something that is closer at hand—perhaps because it is more threatening or more vivid? The proposed explanation is based on the fact that moral norms are themselves high-level constructs. Because of this, a high-level construal of a given situation can more easily be brought under the category of a moral norm, compared to the concrete, low-level construals we make of events that are psychologically closer (Eyal and Liberman 2012, 188–189; Eyal, Liberman, and Trope 2008, 1208, and their Study 1).\(^10\)

Take lying as an example. The concept of a lie can be seen as having a prototypical structure. This means that we have some idea of a prototypical instance of a lie, and that we extend this idea in various directions to incorporate other, less central cases (Johnson 2014, 91–98). If, then, we are confronted with something that might be a lie, and we construe the case in abstract, high-level terms, then aspects of the case that may align it with the prototypical lie—what we take to be the speaker’s intention, for example—will feature prominently in our thinking about it. At the same time, the details of the case will not be considered, details that might have distinguished this case from the prototype and pointed towards other, less morally problematic high-level descriptions, like a white lie or an ironic comment. Thus, high-level construals may make us trigger-happy in our moral judgements.

\(^9\) Specifically, participants read vignettes adapted from Haidt (2001) and Haidt, Koller, and Dias (1993) on such things as eating the family dog, consensual sexual intercourse between siblings, and cleaning a house with the national flag, as well as on a married woman having an affair, and a student cheating on an exam (Studies 1–3). Similar vignettes describing morally virtuous actions were also used (Study 4). Each vignette described what the authors say is a moral violation/morally virtuous action (high-level information) as well as contextual details that moderated the offensiveness/virtuousness of that violation (low-level information). Psychological distance was manipulated either by instructing participants to imagine the events taking place tomorrow/in a year (temporal distance), or from their own perspective/that of a third person (social distance). The effect of stronger moral judgements (negative and positive) with increased psychological distance was consistent across experiments. From the point of view of a moral philosopher, the use of vignettes describing purported ‘moral violations’ that are not necessarily immoral, such as cleaning one’s house with the national flag, is problematic. Fortunately, other studies have replicated the effect using less controversially moral/immoral examples, such as giving to charity/stealing (Agerström and Björklund 2013a, 2013b).

\(^10\) A related explanation is that increased psychological distance leads to increased ‘dispositionalism’; that is, the tendency to ascribe people’s behaviours to traits rather than circumstance, see Agerström and Björklund (2009, Experiment 3).
In general, the greater the psychological distance, the more abstract the construal. The more abstract the construal, the fewer the details that can distinguish the present case from the prototypical moral transgression. The fewer the details that can distinguish it from a prototypical moral transgression, the easier it is to categorise it as a moral transgression. The easier it is to categorise it as a moral transgression, the easier it is to condemn it—possibly in error. Conversely, the smaller the psychological distance, the more the contextual details of the situation occupy our thinking about it, the less likely we are simply to categorise it as an instance of a prototypical moral transgression, and the less likely we are to condemn it, whether rightly or in error. Thus, if these findings hold, it would be an empirical demonstration of the old saying, *tout comprendre rend tres indulgent*.\(^\text{11}\)

These effects have been replicated (Mårtensson 2017 counts 12 congruent findings), but the opposite has also been found (Gong and Medin 2012), complicating the picture. Specifically, Gong and Medin manipulated construal level by asking subjects to consider the “why” (high-level) versus the “how” (low-level) of a moral transgression (2012, 629), and found that moral judgements were stronger for low-level construals (replicated in Žeželj and Jokić 2014). Responding to these findings, Eyal, Liberman, and Trope (2014) tested and found that the why-manipulation induces participants to search for explanations for why someone committed a moral transgression, which they argue serve as mitigating circumstances.

Therefore, the direction of the effect—weakening or strengthening moral judgement—may depend on whether psychological proximity and/or low-level construals lead to emphasis on mitigating or exacerbating circumstances. In Eyal, Liberman, and Trope (2008) the contextual details were designed to be mitigating, but the opposite could of course also be true. Indeed, Williams, Stein, and Galguera (2014) found that greater psychological proximity was correlated with a stronger affective response to the situation. If that response is negative, as it is apt to be in cases of serious moral transgressions, psychological proximity could lead to harsher judgements in spite of the reduced salience of a general moral norm covering the act in question. Hence, even though it may often be true that *tout comprendre rend tres indulgent*...

\(^{11}\) The expression is usually given as *tout comprendre c'est tout pardonner*, but the version from Germaine de Staël’s *Corinme ou l’Italie* (1807, quoted in Thorstensen 2011) fits the context better.
indulgent, “understanding”—as P. F. Strawson remembers J. L. Austin remarking—“might just add contempt to hatred” (Strawson 1985, 39 n5).

The point, however, remains: construal level and psychological distance affect our moral judgements. We judge the same case differently depending on whether we construe it in low-level detail or high-level abstraction, as psychologically proximate or psychologically distant. In other words, the problematic entanglement of affective and physical space extends into psychological space. In a weak sense, the tension between impartiality and understanding is also reproduced. As soon as we bring home to us the case of some particular person, we treat her differently to all those whose cases we have not thus brought home to ourselves; we no longer treat her as “but one in the multitude” (TMS, II.ii.2.1, 97).

There is also a stronger sense in which the tension picked out by Forman is perpetuated in psychological space: four of the studies reviewed by Mårtensson (2017) investigated the effect of social distance on moral judgement. All these, including the only high-power replication study in the field (Žeželj and Jokić 2014), found that social proximity leads to weaker moral judgement. A fifth study not mentioned by Mårtensson may serve as an illustration of the effect. In a series of experiments, Gino and Galinsky (2012) found that inducing perceived social proximity between participants and the people they judged would lead participants both to make more lenient judgements of immoral acts, and be more disposed to acting in an immoral manner themselves. For example, simply getting participants to perspective-take with someone who had acted selfishly (by keeping to themselves a reward that could be split) would lead participants to judge the selfish action less harshly, in turn increasing participants’ own intention to act selfishly (Experiment 1). This is the same kind of indulgence Forman picked out as the threat to impartiality that arises from close physical contact. In other words, the very act of bringing the case home, of imagining ourselves in the place of another, will tend to make us more indulgent towards those we judge.

12 We may also judge groups of people differently, see Luguri, Napier, and Dovido (2012).
13 Interestingly, the effect was less pronounced for generous actions (Experiment 4).
14 Including, it would seem, when we judge ourselves; see Conway and Peetz (2012) for a link between construal level, moral licensing, and moral reinforcement.
There is, possibly, a further, still stronger sense in which psychological proximity perpetuates proximity bias in our moral judgements. For, if becoming well-informed really requires us to bring home to us the case of another in all its minutest incidents, then the brute fact of cognitive limitations—specifically, limitations on the number of “chunks” of information that we can hold in our working memory at any given time (Cowan 2001, 88; Mesoudi, Whiten, and Dunbar 2006; Meyer, Taylor, and Lieberman 2015; Stiller and Dunbar 2007)—seems to entail that we will only be able to imagine, in requisite detail, the situation of a single person at any given time. In a conflict involving two or more people, therefore, it seems we would judge the relative merit of each of their claims differently depending on whose case we last brought home to ourselves. Hence, even if we understand impartiality narrowly as equal consideration of the interests of those directly involved in a case, our ability to judge impartially may be threatened by bias arising from a purely psychological proximity.

This, admittedly, is the most speculative of my proposals, and my overall argument does not depend on it being sound. Neither does my overall argument ride on CLT offering the best description of the effect of psychological proximity on moral judgement. Williams and Bargh (2008), for example, argue that the effect can be explained by how metaphors based on physical space shape our thinking (but see Pashler, Coburn, and Harris 2012 for a replication failure). Yet, the overall effect they aim to explain remains the same: psychological distance and/or construal level affect judgements, including moral judgements. Whether the effect is primarily cognitive or affective, mediated by changes in construal level (low-level, ‘minutest incidents’) or psychological distance (near-social, bring the case home), and one of stronger or weaker—harsher or more lenient—moral judgement, the overall effect appears to be robust. The implication is also the same: taking physical (or any kind

15 Stiller and Dunbar found that people could keep track of five orders of intentionality (including their own) at the same time, which, interestingly, correlates with the typical number of close friends (5). However, the task involved very simple chunks of information. Thus, for the fifth level of intentionality, the sequence to remember ran as follows: “Jenny thought that Emma hoped that the boss would believe that the chemist wanted Emma to work for him” (Stiller and Dunbar 2007, 102). Impressive as this may be, it does not include the kind of ‘minutest incidents’ that Smith wants us to heed in the exercise of our sympathetic imagination.

16 Indeed, this effect is tacitly recognised in the rule, observed in some legal traditions, of letting the defendant have the last word in court (for a paradoxical consequence of this, further underlining the complexity of moral and legal judgements, see Englich, Mussweiler, and Strack 2005).
of actual) distance out of the equation does not eliminate the tension between understanding and impartiality.

VI. IMPLICATIONS

The tension pointed out by Forman is not solved by taking physical distance out of the equation. This presents a problem for those who think we can transcend conventionality by reflecting on the moral norms of our culture from the perspective of an impartial spectator. For, unless we can also somehow take psychological distance out of the equation, it is hard to see how we could resolve the tension and be both impartial and well-informed at the same time. That is not to say that such a resolution is impossible. However, as long as we have not resolved this question, appeals to a spectator who we simply stipulate as being impartial and well-informed will be of no help to us. If we cannot trust ourselves to be both impartial and well-informed, no appeal to the idea of an impartial and well-informed spectator would give us the critical distance we need on our moral culture.

Of course, for those who, like Forman, were already convinced that the impartial spectator could not help us transcend convention, this conclusion might not seem threatening. However, as noted in the introduction, the reasons to want to save the impartial spectator extend beyond this debate. Even among those who think the impartial spectator cannot critique a given conventional standard of propriety, it is commonly accepted that an impartial spectator can criticise the partial or ill-informed application of that standard to a particular situation (Fleischacker 2011, 28–29). Yet, if the above argument holds, it seems we might lose even this. Another way of putting the problem is to say that, if impartiality and understanding stand opposed even in our psychologies, then the ideal of the impartial and well-informed spectator may turn out to be out of reach because it is psychologically unrealistic.

Why, one might ask, is this a problem? After all, an ideal is something we had better not be able to achieve, if we are to have something to strive for. That might have been a fair point, if it were not for the fact that one of the supposed virtues of Smith’s theory, and one of the things that makes TMS relevant today as an alternative to later theoretical developments, is precisely that the ideal of the impartial spectator is supposed to be a realistic, achievable ideal. When Smith first joins the adjective “impartial” to the noun “spectator”, the spectator in question is a real person, someone like you or me (TMS, I.i.5.4, 30). When
Smith moves on to discuss how we judge ourselves (in analogy to how we judge others), the impartial spectator comes to stand for the idea of such a person; someone we can imagine being present, whether or not she actually is. However, even in the shape of this idea—which, subsequently, becomes our ideal—the impartial and well-informed spectator retains its element of realism, remains a point of view we can in principle inhabit, if not always in practice (see for example TMS, III.3.4, 158).

This in contrast to later, more explicitly idealised conceptions of the impartial spectator, such as Roderick Firth’s “ideal observer” (1952, 321). Where Smith says that an impartial spectator must be well-informed, Firth posits that the ideal observer is both omniscient and “omnipercipient”—all-knowing and able to imaginatively visualise anything (Firth 1952, 333–335). Where Smith’s impartial spectator is one who is disinterested or dispassionate in a given situation, Firth’s ideal observer is “entirely lacking in particular interests” and incapable of emotions like jealousy—or even incapable of emotions altogether (Firth 1952, 335–340). In other words, the ideal observer is truly ideal. This is a reflection of the job for which Firth formed it. The ideal observer is meant to provide a theoretical grounding for the possibility of objective moral truth, without the “queer” metaphysics usually involved in realist accounts of moral objectivity (Mackie 1977, 38). The ideal observer is not an actual person, but a purely hypothetical entity (Firth 1952, 325).

Although TMS is sometimes presented as an early version of Ideal Observer Theory (for example by Rawls 1999, 161), Smith scholars tend to reject this view (see for example Broadie 2006, 184; Fleischacker 2016, 274; Frazer 2016; Haakonssen 2002, xviii; Raphael 1975, 94–96; Weinstein 2016, 353; and, more obliquely, Carrasco and Fricke 2016; Sayre-McCord 2010). The impartial spectator, these authors argue, is not some abstract philosopher’s invention, some purely theoretical device around which to build a metaethics of moral truth. The impartial spectator is supposed to be a workable ideal for everyday moral judgement—indeed, if Smith is right, the impartial spectator is our ideal: the ideal we already, if tacitly, endorse and aspire to (TMS, III.3.2-3, 156–157). The conclusion, though often implicit, is that this slightly non-ideal realism also makes the impartial spectator better suited as an ideal for actual, everyday moral judgement compared to the unachievable ideal observer.
Therefore, that pesky second adjective ‘well-informed’ throws a wrench, not into a philosopher's attempt at grounding the possibility of something like objective moral truth (as Firth attempts to do with his ideal observer), but into our actual attempts at being fair and impartial (which is what Smith takes himself to be describing). Granted, the fact that there is a tension between understanding and being impartial is less problematic than the extreme kind of idealism involved in the ideal observer. However, for the purposes of everyday moral deliberations, the two ideals may turn out to be on a par.

There are (at least) three ways to respond to this. The first is to note that psychological studies like those reviewed in this article reveal tendencies, not hard limits on what we can and cannot achieve. Hence, we may be able to control for the psychological tension between understanding and impartiality if only we try.17 This response merits closer attention than I am able to give it here, indeed it merits attention from researchers working within the framework of CLT. However, extant research on our ability to consciously control for bias in our judgements throws a relatively pessimistic light on the prospects of this first response (see for example Lilienfeld, Ammirati, and Landfield 2009). Therefore, it behoves us to consider the two other kinds of responses as well.

The second is to follow the example of Jason Kawall (2006), who, in response to Geoffrey Sayre-McCord (1994), has made a similar argument against David Hume's supposedly non-ideal "general point of view" (Kawall 2006, 370). Kawall argues, 1) that the general point of view is just as psychologically unrealistic as is the ideal observer, but 2) that it does not matter whether our moral ideals are psychologically realistic and, therefore, 3) that the ideal observer is preferable to the general point of view in virtue of its claim to all-seeing, all-knowing universality (which, among other things, would allow it to critique conventional standards of propriety). If I am right, the same argument could be made about Smith's impartial spectator. However, since adherents of non-ideal spectator theories object to 2), Kawall’s alternative is unlikely to be adopted by them.

What about the third response? Well, we could revise the impartial spectator to make it even less of an ideal, or, rather, make it non-ideal in such a way that it again becomes a psychologically realistic ideal. What would it take to make the impartial spectator such an ideal? Answering

---

17 Thanks to an anonymous reviewer for this suggestion.
that question would take an investigation that I cannot begin, let alone finish, within the scope of this article. However, for those who find in TMS a sketch of an adequately explanatory and interestingly justificatory account of moral judgement, it is, I think, an investigation worth pursuing.

VII. CONCLUSION

In an article on the foreign travels of German Members of Parliament, the political scientist Berit Bliesemann de Guevara recounts how one of her informants summed up the problem of proximity: “ignorance protects from granting’, meaning that it is much harder to reject funding for a project once you have been on site, seen the project and met the people behind it” (Guevara 2016, 66).

The above argument amounts to the claim that this is true also for purely imaginary transpositions. Bringing the case home suffices to produce proximity bias in moral judgements. To the extent that becoming well-informed requires us to bring home to us the cases of those we judge, then the tension between understanding and impartiality defies easy resolution—even if we take physical distance out of the equation. The impartial and well-informed spectator remains an elusive ideal.

How we respond to this depends on what we want from our ideal. If we want an ideal that makes sense of how we can escape the confines of moral conventions, and, moreover, is usable for everyday moral judgements, we would do well to reconsider the contours of the impartial and well-informed spectator. How impartial? How informed? Those are questions for future research.

REFERENCES


**Sveinung Sivertsen** is a PhD candidate at the University of Bergen, Department of Philosophy. His research focuses on the moral philosophy of Adam Smith, and in particular on how modern psychological research can cast new light on old problems in Smith’s sentimentalist ethics.

Contact e-mail: <sveinung.sivertsen@uib.no>
(History of) Economic Knowledge Freed from Determinism: An Interview With Joseph Vogl

JOSEPH VOGL (Eggenfelden, 1957) is a prominent German philosopher, working in intellectual and media history. He is Professor of Modern German Literature, Literary, Media and Cultural Studies at the Humboldt University of Berlin and permanent Visiting Professor at the Department of German at Princeton University. Before coming to Berlin, he served as a Professor of Theory and History of Artificial Worlds at the Department of Media studies at Bauhaus University in Weimar. In 1984, he received an MA at the University of Munich after studying philosophy, history and German literature in Munich and Paris. In 1990, he received a PhD in Modern German literature from the University of Munich and in 2001, he obtained a *venia legendi* from the same university.

Vogl's research spans literature, history of knowledge, history and theory of media, and political philosophy. He is the author of *The Place of Violence. Kafka’s Literary Ethics* (1990); *Calculus and Passion: Poetics of Homo Oeconomicus* (2002); *On Tarrying* (2007/2011); *The Specter of Capital* (2010/2014); *The Ascendancy of Finance* (2015/2017); he also edited several influential collections and authored many shorter essays in various fields. Vogl is known, among other things, for the notion of the ‘poetics of knowledge’ and, related to that, for the genealogy of modern economic thinking which he explores both in its political and in its literary form. In recent years, he shifted his attention to the analysis of modern capitalism, all the while continuing his work in the history of literature, political philosophy, and discursive history of danger and risk.

The Erasmus Journal for Philosophy and Economics (EJPE) interviewed Vogl about his intellectual career, his relationship to the history and philosophy of economics, and his perspective on the analysis of contemporary capitalism.

EJPE’S NOTE: This interview was conducted by Ivan Boldyrev. Boldyrev is Assistant Professor at Radboud University Nijmegen, specializing in the history and philosophy of recent economics, history of philosophy, and critical theory. He thanks Ola Morris Innset for suggesting one of the possible questions while preparing the interview, as well as Jakob Schönhuber and Noah Willumsen for translating the text from the German.
EJPE: I suggest we begin from the very beginning. You studied German literature, history, and philosophy. Why did you choose to engage with these disciplines? How did time and place shape your choices in the university?

JOSEPH VOGL: Such decisions lie in youthful semi-darkness, and one must beware of lapses. But probably I was attracted to these subjects by a certain opacity, an inability to fully understand. Unlike in any other discipline, you are not blinded so much by clear subjects and boundaries or subfields. Rather the curricula looked labyrinthine; the straying was just as foreseeable as the strolling around. So, it was probably the unsystematic and the groundless that attracted me to literature, philosophy, or history. Although it could occasionally be frightening. All this happened at the end of the 1970s, and at that time places like universities—at least if you came from small towns in Lower Bavaria—smelled less like education (Ausbildung) than like release into the wild (Auswilderung) and freedom. And once, when I was about to escape from this difficult terrain, an academic teacher stopped me and convinced me to stay.

You started your academic career as an intellectual historian and a scholar of literature. What motivated you to turn to economic knowledge?

That had almost a certain inevitability. When I left for Paris in the early nineties, equipped with Foucault's The Order of Things (1970 [1966]), reading, with hindsight, traces of anthropological knowledge in literature and aesthetics of the 18th century, I came across economic ideas on almost every page, in natural history, in medicine, in philosophy, in encyclopedias, in the theories of signs and in the teachings of beauty. There was circulation, communication, flows of exchange all over the place, too much compensated for too little, providence worked through all the domains of nature, and the Robinsons were the model. This reinforced the impression that the economy of knowledge was reflected in economic knowledge and that the human being, exemplary of its species, was engaged in aligning himself with homo oeconomicus. The question of the status and genesis of economic knowledge has therefore become unavoidable for me.

The risk of simplification notwithstanding, how do you see the guiding idea of your first big book on economic knowledge—Kalkül und
Leidenschaft/Calculus and Passion (2002)—today? Could you also elaborate on how the ideas of that book were connected to your broader project on the ‘poetics of knowledge’?

If one considers such books again retrospectively, one feels the hopeful intellectual rumbling of that time, while the legs still stick to the academic flypaper. One had to prove something not only to others, but also to oneself. And, furthermore, one needed to become recognizable by the so-called ‘scientific community’, the literary studies. For me, the task was to prove that objects and epistemic modes of the emerging economic knowledge not only became a dominant pattern for the description of natural relations, moral doctrines, and forms of social communication, but also produced privileged modes of representation, whose program could then be traced in narrations, plays or aesthetic concepts. Every epistemological clarification, as I said at the time, is linked to an aesthetic decision; every knowledge system develops options for representation that determine the consistency and correlation of its objects. This resulted in the ‘poetics’ of the type called homo oeconomicus. I examined its actions, impulses, and intricacies across different discourses and genres: in theatre, for example, they were shaped by the blueprint of exchange, money circulation, and contract, whereas in a novel, they were formed by the challenge of overcoming the contingent mass of events, which was also the concern of probabilistics or cameralist, policy-wissenschaft manuals. Ultimately, it was about the logic of a glove turned twice upside down: how we can relate the epistemic substrate of poetic genres to the ‘poetic’ constitution of the forms of knowledge.

Did you ever feel—academically, intellectually—that you belong or belonged to a particular ‘school’ or group or ‘generation’?

No, I never had that pleasure—or that bad luck. By the way, people had tried repeatedly to make a ‘generation’ out of my generation—for example, the ‘1978s’, analogous to the ‘1968s’—as if something would be missing if one could not refine the biological origin with a historical or cultural one. These attempts have been thoroughly unsuccessful. The only intellectual link that remained was probably that we are among the baby boomers and will pose real problems for the pension funds. However, I was looking for alliances, something like food supplements or vitamins, that would make intellectual progress easier. And, given that they offered the least resistance, I turned first to the dead, Adorno
and Benjamin. They had left behind the monuments, the *Aesthetic
could be understood or misunderstood so productively that they gave rise to
something like a thinking feeling of oneself *[denkerisches Selbstgefühl]*, a
felt self-thought. Such texts also worked like trampolines, enabling one
to jump higher than one could with one's own forces, combined with the
impression that, at least for a few moments, one could grasp the
situation of the world.

*Apart from Foucault (and I hope to come back to his work), which
authors or approaches did you find most inspiring in the 1980s and
1990s?*

Apart from the critical theorists and later Foucault, this was initially
Robert Musil’s *The Man Without Qualities* (1995 [1943]) in the late
seventies. The first real reading: seven days in a room, a chair moved to
the middle, disciplined reading posture, following the maelstrom of the
first thousand pages. I couldn’t believe that something like this existed:
a novel dismantling narratively the twentieth century with all its hopes,
stupidities, and ideologies. My best friend, Roger Willemsen, called this a
'senti-mental' project, a stimulation of intelligent feelings and affectively
sharpened thought movements. I remember being excited about the
discovery that literature can be an instrument of knowledge, and even
more, by the related call to apply accuracy and analysis to life and moral
matters. Reading has had an impact on life, and it was no longer
possible to carry on as before. This has probably made me an
idiosyncratic reader with a penchant for idiosyncratic texts, that is, texts
that initially provide less answers to questions than questions and
problems to existing answers. And perhaps that was also the bridge to
Gilles Deleuze (and Félix Guattari)—the *Anti-Oedipus* (1977 [1972])
already circulated outside the seminar rooms at that time and attracted
people like me with its reputation of a wicked work. I had hardly
understood anything, but despite—or because of—this made a
pilgrimage to Paris at the beginning of the eighties, like many others.
Anyway, I stayed faithful, heard Deleuze talking about cinema, I could
follow him well despite my poor language skills, and then, since the
beginning of the nineties, I have acquired the privilege of translating a
few of his books. Basically, it was a stroke of luck and the right course:
the encounter with a way of thinking and an intellectual pedagogy that
made connections to other attractions—to cinema, to Nietzsche, and (in particular) to Kafka.

*The analogy between economics and literature is often drawn by referring to the idea of fiction—embracing both theoretical economic models and fictional narratives circulating on the markets. Do you think this analogy is helpful today?*  
The talk about fictions in the economy has created many misunderstandings, especially the misconception about the contrast between a so-called real economy and the excesses of the ‘fictitious’ financial sector—as if capital accumulation and financial capital were not the engine of capitalist economic activity. But, of course, economic doctrines have never done without narratives, legends, or pictorial ideas, be it the stories of the hard-working and pious Robinson on his lonely island, of the good and faithful merchant or of the evil usurer, be it the image of that fabulous ‘invisible hand’ that is supposed to turn all business malignancies into good. While economics had little interest in the self-enlightenment of its own fictions, literature and aesthetic sensibility on the contrary, were repeatedly drawn to the miracle of economic activity. Thus, the incomprehensible riches of early trade capitalism were reproduced in the magic of an inexhaustible ‘bag of fortune’ in the first German prose novel, the *Fortunatus* in 1509. And it were precisely the workings of the stock market, of the financial markets, have proved, with their adventurous aspects, to be a major challenge for literature—perhaps because in the turbulences there, it is not so much the facts that count as the expectation of facts. Thus one has seen a certain affinity in the speculative spirit and in the play with the non-existent, as Goethe, who once made his Faust—with Mephisto’s fierce participation—print paper money, equipped the whole with poetic wings, with fluttering, airy, unearthly movements, and recognized in it a proximity to the genius of poetry. Similarly, Émile Zola speaks of the stock exchange speculator as a poet of sublime sums of money—and in Don DeLillo’s *Cosmopolis* (2003), the speculator’s mobile office is ‘proustified’, that is, sealed with cork against the rest of the world, like Proust’s writing room. An aesthetic universe. In general, an aesthetic of the sublime comes into play again and again: in view of unimaginable sums of money, to which no intuitive counterpart is available and which enables one to abandon the sluggishness of the material world, the realm of bodies, and use values. In such contexts, Marx identified the
process of capitalization with the formation of ‘fictitious capital’; and perhaps economics is precisely the discipline most saliently demonstrating the potency, the efficiency of fictions in general (as opposed to the creations of fantasy). There, mere expectations, non-existent things and imagined futures directly produce system-level effects. So yes, there are analogies between literature, aesthetics, and economy, but they lie above all in the question of what one can do with fictions and signs. As Mallarmé once wrote: "[t]out se résume dans l'Esthétique et l'Économie politique" (Mallarmé 1895, 79): everything boils down to aesthetics and political economy.

**Do I understand correctly that it was the reading of Musil and Foucault that initially motivated you not to discriminate between reading prose, reading economic texts, and, in a certain sense, reading the economy itself?**

No. It makes a great difference to read literature, and it would be crazy to confuse novels with scientific treatises. But I think I’ve learned two things. On the one hand, it seems worth trying to read treatises (such as economic ones) not as literature, but rather *like* literature. In other words, to pay attention to all procedures—the rhetorical, the medial, the institutional—that are involved in the production of ‘truths’. This is the only way to grasp the historical singularity characterizing the systems of knowledge. There are no ‘events’ or ‘references’ waiting in a timeless and unmoved outside to be awakened and made visible by discourses, by statements, by the propositions of existence. Every characteristic, every conceptualization of an epistemic object simultaneously carries out a discursive realization of the same object, a manufacturing in which the codes and values of a culture, the systematics and practices of an epistemic field are reproduced. On the other hand, it may be helpful to not always read literary texts only as mysterious documents that are in need of commentaries and interpretations in order to finally be better—or actually—understood. One could reverse the perspective: literary texts themselves are interpretations and readings, and literary history could be a history of different interpretive techniques represented by the texts themselves. Thus, there would be two hermeneutical perspectives, leading us in opposite directions: one ‘theological’, as it were, in which past texts wait for a kind of redemption by the present interpreter; and one closer to a ‘materialistic’ kind, with which one looks at the peculiar performances and powers of
the past productions of meaning. This may also give us the opportunity to read literary and, for example, economic texts together without ignoring their differences.

In Germany, the history of economic ideas—still labeled, somewhat embarrassingly, Dogmengeschichte—is not properly institutionalized and exists at the margins of many different disciplines. But taken as a whole, it is an overlap of different vibrant communities, with various sensibilities, specializations, preoccupations, etc. When you were writing Kalkül und Leidenschaft and Das Gespenst des Kapitals/The Spirit of Capital (2010), did you feel (or assume) that you also address these people—the historians, sociologists, or philosophers engaging with, say, David Hume, Vilfredo Pareto, or Herbert Simon? How would you describe your relation to—and your travels through—this literature?

I have not thought of such readers, of the various experts on different disciplines. Had I done so, I would have lost my courage. For the expert naturally possesses an evil and merciless perspective on the amateurs. At their core, these books of mine were more of a self-experiment: by what means, sources, forms of writing and techniques of reading can I convince myself that some of my speculations and conjectures—on economic contexts, on the status and form of economic theories—are correct? Only in this respect was there an imaginary but completely unspecified audience. Self-conviction only works if one imagines reading eyes behind one’s own back that are not one’s own, but somehow sympathetic. There is always more than enough self-criticism, but the imagined and benevolent reading eyes serve to loosen the blockades of self-criticism. So, I have tried an egoistic approach to the experts in the history of economic thought—that is, to the exponents of the knowledge I am unfamiliar with: I read all their stuff and put them ready like ammunition, but avoid imagining them as readers of my own work.

And related to that: do your regimes of reading change when you switch from the older political economy to today’s economics? In other words, do you believe that after so many transformations since the Enlightenment, the discussion of economic knowledge requires a radically different approach, a new hermeneutics?

The reading of economic texts always went in two directions, forward and backward. Thus, the challenge with the 18th-century texts—from
cameralists to physiocrats to English liberals—was not to read them as old and outdated acquaintances. For when one looked more closely, those texts became vibrant, they were full of excitement about all kinds of novelties: an economic science did not yet exist, one experimented with a hodgepodge of different practical, legal, theological, anthropological, commercial, and political knowledge, invented the market mechanisms as Columbus’s egg in social theory. It was simply an adventurous knowledge having the strangest exponents—writers, moral philosophers, lawyers, secretaries, merchants, schemers, gamblers, charlatans… It was about the cumbersome birth of knowledge, about the birth of facts that were not there before. The current exponents of the economics profession are to be treated in the opposite way. The historical perspective, I believe, was helpful in alienating them a little. Is it really so self-evident to speak of ‘the’ economy and its ‘laws’? What happens when one reads the claims and theories of contemporary macroeconomics as historical documents, as artifacts from the past? And what if we introduce the difference of the historical into the present? This knowledge is thus un-disciplined, it loses its secure academic, institutional, dogmatic anchors and becomes accessible for dispute, for political debate. What interested me about contemporary economics was the following question: How could there emerge a specific interpretation of the world in such a way that the world can be programmed according to this interpretation? How does one deal with a science that deals with interpreting the states of affairs it has created itself? How does one interpret knowledge that is committed to the implementation of its interpretations and construals? That was one of the hermeneutical questions I was interested in.

*The type of reading you propose is provocative. But don’t we then miss an opportunity to grasp the particularity of the present moment and to trace the qualitative changes experienced by today’s market economies (like financialization)? Or, do you believe that the opposition should be elsewhere, not between the foregone past and the radically new present? What would then be the best way to conceptualize the change of capitalist economic order?*

It is probably worthwhile to develop a certain sensitivity for different temporal durations, an intuitive understanding that we do not drift along a well-ordered, consistent stream of time, but exist simultaneously in several layers of time. This applies especially to the contemporary
capitalist market economy. Some inventions and realities that date back to the early modern era are still valid today, they are thus recent: joint stock companies, various forms of enterprise, business models, the ideas of profit, trading on the stock exchange, accumulation of capital, utilization or exploitation of labour—in this respect, the stories of capitalism, as they have been told since Marx, are news from the past that does not want to pass away. On the other hand, it is precisely these powers of perseverance and their 'long duration' that themselves—as Marx and Engels put it—make history; allowing all that is solid ('Ständische und Stehende') to melt into air, undermining certainties, setting a pace of modernization and innovation, and even producing something like a 'journey of change' that can be felt by everyone. What great dynasties and territorial states have failed to produce since the 16th century, namely true 'world empires', was then accomplished by a capitalist world economy that justified the title of a ‘world system’ (according to Wallerstein and others). The interplay of novelty and the ever-sameness thus makes this historical experience so special and makes one suspicious towards the noise produced by the 'latest rage'.

I would direct the sense of history, of the historical atmosphere less to the meaning of the newest and the topical, but to something else: to the question of how the history of capitalist economy can be understood as a change in power and government technologies. The example of ‘financialization’: this is not merely characterized by the take-off of the financial markets, by new financial instruments, or by the rebuilding of corporate structures. Rather, financialization implies that the reproduction of financial capital dictates all other—social, economic—reproduction processes. In other words: only by interpreting economy, market, capital as political forces, as instruments of governmental power we can observe the breaks and shifts in the capitalist economic order, including current dependence of all domains of life on the financial markets.

Would you agree that it was this ‘ politicization’ of economics you've just mentioned, recognizable in the many ways it now infiltrates political agendas, that makes economic knowledge somewhat more ‘vulnerable’ and potentially more accountable (I am following here the remarks of Fourcade, Ollion, and Algan (2015))? Or do you see other consequences of this process?
Two issues about this ‘ politicization’. At stake here, of course, is the supremacy of economics in the social sciences, a dominance that is accompanied by the desire to be 'pure' or 'objective' science like physics, and that carves out the direct access to the ruler by raising a claim to 'fortune-telling'. Every government has economic ‘experts’ belonging to its closest circles—and it's no coincidence that in Germany they have been given, quite seriously, the title of Wirtschaftsweisen (‘economic sages’)—and setting the direction for economic policy. Crucially, this involves the reference to the so-called market laws. The focus on spontaneous and quasi-natural market regimes (Marktordnungen) has obscured the idea that economic knowledge is always about which interests have priority—and for what reasons, despite what kind of resistance. ‘Politicalization’ would thus be linked to the withdrawal of the presumed innocence against political decisions based on seemingly compelling market mechanisms. Economic, financial, and fiscal policy is not dispensed with procedural responsibility; and the usual recourse to those alleged determinisms, which assume, for example, a legal connection between the share of public spending and growth, price level and money supply, competition and the public welfare, salary sacrifice and employment, only conceals the power of the decision-making potential embodied therein. Politicization here means distancing oneself from the alleged laws of the market and making economics appear as the contested field which it has always been and in which forms of economic research cannot be separated from the definition of political goals or preferences. It should be remembered that economics provides a bundle of conflicting interpretations of social and historical facts. On the other hand, as already indicated, one could think of the genealogical or ‘governmental’ perspective proposed by Michel Foucault. From this perspective, economics has emerged since the seventeenth century as specific knowledge of governmentality—close to the ‘reason of state’ idea—linked to the question of how territories, populations, and commerce between people and things can be governed or controlled in a better, safer, smoother, and more efficient way. Thus, the emergence of economic science would always be interwoven with a technical guide to governance, or a set of political guidelines.

What role did the crisis of 2008 play in changing (or reinforcing) your intellectual priorities? Was the book on capital already in the making or was it a reaction to what was happening?
The starting point was, apparently, a threefold astonishment. First of all, the astonishment on the part of so many in the economics profession that the crash of 2008 could actually happen. Hardly anyone had anticipated it, and people even spoke of an intellectual catastrophe for the economic science. Then there was the astonishment at astonished experts, because basically similar financial crashes had been happening routinely since the end of the eighties. Finally, the astonishment that the prevailing wisdom of financial economics survived, almost unscathed, the global financial and economic crisis. This was already apparent in 2009, when I began to work on The Specter of Capital (2014). Actually, I asked myself a question from Voltaire’s Candide (1759): What if we conceived of the 2008 financial earthquake in the same way as of the Lisbon earthquake of 1755? At that time, the previous attempts at proving the existence of God and constructing theodicy became obsolete and survived only as satirical, ‘panglossian’ figures. Wouldn’t something similar be conceivable for those economic and financial theories that fabricate the best of all economic worlds, such as markets full of equilibrating mechanisms and harmonies? What would a financial-economic satire look like? When does science become involuntarily ridiculous? In retrospect, I have tried to find Dr. Pangloss—in all his variations—in economic dogmas, or, to put it differently, someone who preaches optimism and determinism, but also, from time to time, disappears into the hedges with the maid.

In The Ascendancy of Finance (2017), you define sovereignty (ascribed to a ‘collective capitalist’), a concept from classical political philosophy, as the ability to transform one’s own risks in the ‘dangers’ for others, to free oneself from the debt and to make oneself into the lender of last resort. But, is there really a lender of last resort in the world of modern markets, where every agent is connected to and dependent upon everyone else? Or, to put it differently, what kind of identity could this ‘absolute’ sovereign possess?

Questions about the place and status of political sovereignty have once again become particularly acute in the recent years; they have decided electoral campaigns, mobilized new nationalisms, and brought variations on Ubu Roi into presidential office. This is likely connected to the fact that, against the backdrop of what is often called globalization, international treaties and commitments, many have wanted to perceive state sovereignty as a mere remainder, as ‘limited’, ‘shared’, ‘pooled’,...
fragmented’, ‘eroded’, or even ‘eliminated’. That’s not entirely wrong, but it should remind us that it was always only on paper that sovereignty was monolithic, compact, and thus truly ‘sovereign’, in political theology, in theories from Bodin to Carl Schmitt. On the one hand, the sovereign was seen as a kind of highest or final creditor; on the other, already in the early modern period, the treasury, the mint, and the financial system emerged as a domain that opened a kind of trading zone between state power and private persons, one characterized by unclear jurisdictions, particular interests and considerations, by trade in offices and privileges, by intermediaries of all kinds, by rent holders, agents, and trustees. Later on attempts were made to introduce order into these complexities through the establishment of central banks, which initially were given the task of financing the state and later received their own sovereign rights, such as a monopoly of coinage, issue of bank notes, procurement of liquidity, and supervision of the circulating money supply. Under the aegis of the ‘financialization’ of the world economy, those prerogatives—such as the control of liquidity and the money supply—have finally migrated from states and central banks to financial markets. What we are dealing with today is the transition from a ‘government-driven’ to a ‘market-driven’ financial system. This has two consequences. First, central banks no longer have an overview of systemic risks or control over the money supply in circulation. Money creation happens in the markets, the latest financial instruments have erased the difference between money and financial assets, and the idea of a determinate and determinable quantity of money must now appear as a historical curiosity. Even with vast amounts of cheap money, the ECB, for example, has struggled to combat deflationary tendencies in recent years. Second, financial markets have become a ‘prison’ for governments, national economies, and societies. That’s what I mean by the role of the collective capitalist, who is embodied in investment and financial capital: using the threat of capital flight, interest rate disadvantages, and reduced investment, the financial markets themselves have become a ‘monetative’ governmental power, a creditor of last resort.

Although your work resists compartmentalization, how would you characterize its place between a critical history of economic knowledge and practices (which would cover your The Ascendancy of Finance as well) and a political philosophy exploring the forms of
today's capitalism? Which of its features or arguments—if any—distinguish your work from a counter-narrative, in which 'history' of economic discourses would become a ‘theory’ of modern economy contesting these discourses?

I would like to translate the last question back into the indicative and repeat it with my endorsement and thanks: yes, I was pursuing a counter-narrative in which a history of economic knowledge would provide elements for a theory that could be used as a critical tool against this very same knowledge and the conditions it creates. That's why, for lack of a better term, I called *The Ascendancy of Finance* a ‘historical-speculative essay’. It rests on the assumption, firstly, that the objects of history, such as forms of economic knowledge, cannot be grasped with a robust or ruthless, that is to say, a predeterminate theory or method. Secondly, that a ‘theory’ of such objects must be capable of making their concrete historical site visible and thus has only local range. And thirdly, that this site, in the case of the economy, consists in the negotiation of questions of power. Against this backdrop I would understand my essay as an essay in (lower-case) critical theory, if critique means the critique of power: in other words an examination of those procedures that enable and influence events, that can channel forms of behavior and expression. There were probably two basic guideposts. One was set by a 'critique of political economy' in Marx's sense, if one subtracts from it the dialectical theory of the 'negation of the negation' as a natural law in the historical process, the other by Foucault's studies on 'governmentality', although one must thereby assume the inconvenience of dispensing with a unified, comprehensively applicable 'theory of power'. Specifically in *The Ascendancy of Finance* a further difficulty arose: in bringing the history of the financial system or of the financial regime into view (which, for me, came into being within a zone of indifference between political practices and economic processes), canonical theories of politics or of the economy, which mostly assume the differentiation of both systems, were not particularly helpful.

Related to that: How easy or difficult is it for you to weave research cultures—media theory, intellectual history, economics—which seem to follow very different epistemologies? Have you witnessed—and/or do you foresee—any productive entanglements or disagreements among these and other disciplines you are communicating with?
Apart from the fact that research cultures, even when they are new and refreshing, have to stabilize themselves very quickly in academia—by demarcations, definition of disciplines, reproduction of support, methodical debates, continual and dense communications, formation of centers, and gestures of distinction—some ‘stylistic’ differences can be recognized across disciplines and research cultures, different styles of thought, with which we can draw cross-sections. One difference was noted by the neo-Kantian Wilhelm Windelband (1904), who distinguished ‘nomothetic’ sciences (oriented toward the search for general laws), from ‘idiographic’ ones, which refer to the singularity of their objects. The other distinction comes from Deleuze and Guattari (1987 [1980]), who oppose ‘deterministic’ types of science to ‘nomadic’ sciences. If we treat these distinctions somewhat freely (separating them from their philosophical-historical contexts), they offer perspectives and instruments with which one can muster commonalities across various territorial claims and enclaves, an epistemological trace in which, for instance, questions of media theory, history of discourses, and history of science meet. Common to all these perspectives would be the question of how to disentangle ourselves from determinisms in the description of historical processes. It’s quite astonishing that economics, with all its affinity for physics, has never been particularly interested in the physics of ordered structures that arise from completely unexpected forms of behavior outside of equilibrium states. The ‘dissipative structures’ of the physical chemist Ilya Prigogine (1977) would be one example: structures that arise from contingent circumstances and then develop into a system or disappear again. In short: for me this is part and parcel of the task of conceiving the singularity of (historical) objects in such a way that they become discernible in their formation, in their becoming, in their dynamic and contingent aspects. That’s very ambitious; but so was the question.

**So, do you believe that a new economic theory of capitalism is necessary? And can we see the contours of this new theory emerging?**

I would prefer the term ‘analysis’ to ‘theory’. Because unlike conceptions of theory, ‘analysis’ or ‘analytics’ doesn’t suggest a sense of closure, even a provisional one: the work of analysis is never done, phenomena are always welling up, you can’t leave the construction site. That doesn’t mean that a theoretical, that is to say, a conceptual survey of contemporary capitalism would be pointless, or that we don’t need to
invent new concepts for new conditions, for example, for ‘financial market capitalism’, for ‘platform capitalism’, for ‘authoritarian’ capitalism, for the ‘linguistic’ capitalism of search engines, for the new *dramatis personae* of ‘influencers’, and so on. But let’s keep in mind: the archives and libraries are full of helpful analyses and theories, there’s no reason to eschew an open-minded eclecticism. Perhaps there are two general criteria for the theoretical orientation of analyses of capitalism, namely to seek procedures of representation that make the situation seem firstly somewhat less tolerable and secondly susceptible to change. To give one example: if you read Hyman Minsky’s (1994) reflections on the instability of financial markets, then the conclusion is not that everything must be done to stabilize them—they are structurally unstable and the next crash is coming one way or another—but rather, that the dependence of national economies, states, societies, lives on financial markets should be reduced. And yes, this opens up new perspectives and disciplines, which are already identifiable. Financial market capitalism is not an economic system, but rather a form of global governance that is creating its own rules, laws, and institutions, detaching itself from territories and nation-states, transforming the geopolitical order into a geoeconomic one, establishing centers of accumulation and zones of exploitation, and immunizing itself against increasingly irrelevant popular sovereignties. Those are new lines of conflict and thus attractors for theory.

*Do you see any connection between the ways economists think about time and uncertainty and the ways time becomes what you call a temporal ‘resource’ of sovereignty, the stuff being traded with, controlled, distributed, and so on?*

For as long as people have been thinking about economic matters, time has been a crucial element. For Aristotle, for example, the natural, cyclical time of becoming and decay was ‘perverted’ by the time of the money trade, in other words, by a time that, through the making of money, interest, and interest on interest, had come out of joint, out of its curve, out of its subordination to the cycle of nature. And in scholasticism, not least among the justifications for prohibitions on usury was the idea that the time of interest accumulation was in competition with the time of creation, the special property of God.

Today the way time is handled could be described as follows: if financial transactions mean trade in uncertainty and risk, and thus
commerce with time, that is, with investments, advances, and forecasts, then recent financing techniques in particular have laid claim to a highly extravagant management of time. This applies above all to what is referred to as securitization or derivatives trading, perfectly financial transactions in other words, in which risks are insured with risks, that is, outsourced, spread, and distributed across different timescales. Price risks are ‘hedged’ by spreading price risks, speculative transactions with speculative transactions, new risk markets are created to collateralize risks, current risks are offset with futures risks, and these with the risk of future futures. Current markets are moved by the after-effects of these respective futures and time is presupposed as an infinite and inexhaustible resource. Accordingly, in the associated theories and models of finance economics, future probabilities are calculated along the lines of past probabilities. Time, that grim and mercurial reaper, seems tamed, the end of history programmed. One might well see in this the desire of capital for eternal life. The sovereign is he who conquers time.

And yet, there is no way to prevent time periods from being finite, deadlines from arriving, payments from coming due. Future presents do not necessarily correspond to present futures, and what we call a crash or crisis is the irruption of finitude: the future has simply gotten too expensive, the resources of time have been used up. At this point the power differential, the role of capital-time finally becomes concrete. For if questions of justice today hinge on questions about the social distribution of economic risks, then a fateful asymmetry can be observed here. While (as after the last crash) the responsible parties and decision-makers in this sector, that is to say the management of big banking houses and investment banks, had to assume no liability whatsoever for the risks they ran, others—homeowners in the United States, retirees in Greece, and so on — were liable for risks that they had not taken on. And so for them, the system has proved to be dangerous. In his last book, *Skin in the Game* (2018), Nassim Taleb described these sorts of asymmetries: especially in the financial system, but also in big business in general, incentives (like bonuses) are given to take on risks, whose consequences are born entirely by others. By this point, at the very latest, in the trade with risks, uncertainty, and time, a tyrannical cynicism has become manifest within the system.
There is now a huge and ever-growing literature—historical, critical, and historico-critical—on the concept and practices of ‘neoliberalism’. What is your attitude to this concept (and has it changed)? How would you define neoliberalism? What do you find useful or helpful in this literature and why?

Everything there is to say about it seems to have already been said. Thorstein Veblen (1900) had spoken of the ‘teleological metaphysics’ of liberal economic theory, Alexander Rüstow (2004) saw in liberalism the continuation of a ‘deistic theology’—that is, of a doctrine that had begun as propaganda for civil liberties and then degenerated into market idealism. Modern liberalism likely emerged out of a conglomeration of moral teachings, political theory, market ideology, and social technology. And today’s so-called neo-liberalism has radicalized these things since the 1970s: the spread of competition across the social fabric, the implementation of micromarkets in all niches of life, the eulogies to the entrepreneurial self, the assertion of formal equality to justify material inequality. Then neoliberalism got bogged down in a political antinomy that made it politically unattractive: the juxtaposition of state and politics on the one side, and market and economy on the other. The same arguments are repeated endlessly, with different conclusions. The forces of the market are invoked to limit the power of the state, the strong state is invoked to fight market excesses. What gets overlooked in the process is that liberalism in its various forms was always interested in the refinement and proliferation of mechanisms of control and rule: what is now called ‘governance’ and is reflected in, for instance, the multiplication of public-private partnerships, is one of the effects thereof. Presumably, the thinking of liberalism has become so unavoidable because it has been realized to a point of discernibility in legislation and institutions, in business models and academic disciplines—a somewhat incoherent but efficient theoretical-practical construct that provides an essential contribution to stabilizing capitalist economic modes. And therefore it can probably be said that neoliberalism today has provided so many answers that it has run out of questions, or more precisely: that it repeats answers to questions that it is no longer capable of posing.

Given this all and coming back to the status and the critique of economics (which is quite frequent in your work): the standard defense strategies of economists after the crisis of 2008 imply that
economic science today is as ‘realistic’ as is possible; that economists have abandoned the pretense of explaining the world as a whole and concentrate instead on small-scale improvements and piecemeal modeling using the most advanced—often (quasi)-experimental—empirical techniques; that, finally, economics is very much like medicine: it cannot make people healthy, but can initiate small, moderate improvements. What do you think about this defense and where would you locate economic science today?

First of all, it was quite conspicuous how current market models and their prognostic power were justified after 2008. A few examples: Ben Bernanke, the former chairman of the US Federal Reserve Bank, believed that the crash of 2008 was the fault of irresponsible actors and did not “require us to rethink economics and finance from the ground up” (Bernanke 2010, 2). The Nobel Memorial Prize laureate Robert Lucas put it even more pointedly: our models and simulations do not provide information about potential crises, only “a forecast of what could be expected conditional on a crisis not occurring” (Lucas 2009). And in another expert opinion it was simply stated that before 2008 pessimistic views were a minority position among economists and therefore could not claim any special credibility. The economic historian Philip Mirowski has spoken of a ‘cognitive dissonance’ in this regard, in other words, of the problem that there is fundamental contradiction between situations and convictions that cannot be resolved within prevailing economic dogma. Even if there is a bit more modesty today, we should not forget that economic forecasting takes pride of place within economic science and modeling. That’s no less true for the latest paradigms, like microeconomics and behavioral economics. They have given up certain oversimplifications like ‘rational choice’, but they still assume that market processes follow some sort of laws and that markets are more or less efficient. I would even agree with Friedrich Hayek here, who once questioned the epistemological status of competitive market models and immediately concluded that the validity of the competition theory “can never be empirically verified for those cases in which it is of interest” (Hayek 2002 [1968], 10, emphasis FH). There are always idealized abstractions at play, inevitably. And that means: only if economic theory reflects upon its own limitations and those of its models, in other words, on the significance of incomplete knowledge, uncertainty and unpredictability, of open, unforeseeable futures, only if
it opens its preserves to the air of history, does it deserves the title of a theory.

REFERENCES
Mr. le docteur Ralph [Voltaire]. 1759. Candide, Ou L’Optimisme. Genève: Cramer.
Reflections on the 2018 Nobel Memorial Prize Awarded to William Nordhaus

J. PAUL KELLEHER

University of Wisconsin-Madison

A definition is just a definition, but when the *definiendum* is a word already in common use with highly favorable connotations, it is clear that we are really trying to be persuasive; we are implicitly recommending the achievement of optimal states.

— Arrow (1963, 942)

William D. Nordhaus of Yale University was one of two recipients of the 2018 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. He shared the prize with Paul M. Romer. In its Scientific Background essay, the Prize Committee lauded each awardee's contributions to the analysis of market failure, and to the development and extension of neoclassical growth theory, for which Robert Solow won the same prize in 1987 and whose roots trace back to a landmark 1928 paper by Frank Ramsey.¹ Whereas Romer's work explores positive externalities in the form of knowledge spillovers, Nordhaus has focused on the negative externalities associated with greenhouse gas emissions. He is credited with developing the first serious Integrated Assessment Model (IAM), a tool for understanding the complex interactions between the climate system and the global economy. Nordhaus' IAMs are also frequently used to evaluate climate policies—“to rank different policies according to their desirability” and “to devise the right dose of the right medicine”, in the words of the Prize Committee (Royal Swedish Academy of Sciences 2018, 26, 2).

Although there has been much debate over the empirical assumptions underlying Nordhaus' analyses,² the primary topic of

¹ Strangely, the Scientific Background paper's bibliography lists F.P. Ramsey as 'A.S. Ramsey' (Royal Swedish Academy of Sciences 2018, 48).
² For example, Dietz and Stern (2015, 578) write: “[W]e assume that the damage function linking the increase in global mean temperature with the instantaneous reduction in output is highly convex at some temperature. Consideration of some of
interest to those in my discipline—moral philosophy—has been Nordhaus' views on discounting the future benefits of greenhouse gas mitigation. His discount rates are among the highest in the climate economics literature, and this is a key reason why Nordhaus has always labeled as ‘sub-optimal’ the popular precautionary proposal to limit average global temperature increases to no more than 2 degrees Celsius above pre-industrial averages. I know of no moral philosopher who accepts Nordhaus’ approach to discounting, and many economists also believe that it violates basic ethical principles of intergenerational fairness and impartiality.

My own view is that Nordhaus' approach to discounting is not necessarily at odds with the approach favored by moral philosophers. What is problematic is Nordhaus’ insistence that his discounting method sheds privileged light on so-called ‘optimal’ climate policy. As I seek to explain in what follows, Nordhaus’ approach to climate policy evaluation can shed light on one—but only one—dimension of the climate change problem. In the normal English language sense of the term, there is nothing particularly optimal about the temperature increases associated with his modeling choices.

I. NORDHAUS’ APPROACH TO CLIMATE POLICY EVALUATION

Unregulated greenhouse gas emissions are a negative externality because those who emit need not take into account the cost they impose on uninvolved third-parties. More specifically, carbon dioxide is a stock externality in the sense that the main driver of the adverse external costs of climate change is the cumulative total amount of carbon in the atmosphere. Because this cumulative total is not something that any individual country can meaningfully influence on its own, countries may see no compelling reason to reduce their emissions today, even if they can foresee that runaway climate change will be bad for them in the far future. As Nordhaus and his co-author Zili Yang wrote in 1996:

---

3 In his most recent book, Climate Casino, Nordhaus says that a 2.8 degree rise is optimal (Nordhaus 2013a, 212). Bizarrely, a slide in his Prize acceptance speech reports an increase of 4 degrees as optimal (Nordhaus 2018, 6). In Climate Casino, Nordhaus reports that his model treats a rise of 4 degrees as optimal given the assumption of low participation among the countries of the world, which increases marginal abatement costs.

---

the science, for example, on tipping points, leads us in this direction. By contrast, most existing IAM studies assume very modest curvature of the damage function. [Nordhaus'] DICE [model's] default is quadratic and it is well known that with the standard values of the functions’ coefficients an implausible 18°C or so of warming is required in order to reduce global output by 50%".

---
The results of this new integrated model of climate and the economy emphasizes the implications of the fact that while climate change is a global externality, the decision makers are national and relatively small. These inherent difficulties involved in planning over a horizon of a century or more about so uncertain and complex a phenomenon are compounded by the dispersed nature of the decisions and the strong tendency for free-riding by nonparticipants in any global agreement. Countries may therefore be triply persuaded not to undertake costly efforts today—first because the benefits are so conjectural, secondly because they occur so far in the future, and third because no individual country can have a significant impact upon the pace of global warming. (Nordhaus and Yang 1996, 762)

Nordhaus views his task as discovering how much greenhouse gas mitigation would be undertaken if countries could overcome the three obstacles that, in the world as it is, ‘triply persuade’ them to emit profligately. For this reason, he refers to his Integrated Assessment Models as “positive”, rather than normative, models. As he puts it:

[IAMs] can be interpreted in two ways: they can be seen both, from a positive point of view, as a means of simulating the behavior of a system of competitive markets and, from a normative point of view, as a possible approach to comparing the impact of alternative paths or policies on economic welfare. (Nordhaus 2013b, 1081)

Yet despite insisting that his approach to climate policy analysis is positive rather than normative, Nordhaus continually refers to the emissions abatement path picked out by his positive analysis as the “optimal” path (Nordhaus 2008, 14, 68), and calls policies that implement that path “idealized” and “economically desirable” policies (Nordhaus 2008, 14; Nordhaus 2013a, 76). So at the very least, the rhetorical use to which Nordhaus puts his results is normative, and not merely positive or descriptive. Yet this indisputable fact—which I discuss further below—should not blur the point that, for Nordhaus, the path he recommends is the path that countries themselves would adopt, in light of their current policy preferences, if they could overcome collective action problems.

Nordhaus’ approach to discount rates is informed by this focus on what countries themselves would do if they were fully informed about the future costs of climate change and could overcome the impulse to
free-ride on others’ mitigation efforts. He calls his approach “the opportunity cost approach” to discounting (Nordhaus 2013a, 188):

Countries have a range of possible investments: homes, education, preventive health care, carbon reduction, and investing abroad. Particularly in a period of tight government budgets and financial constraints, the yields on such investments might be very high. [...] A country would be poorly served to put its scarce funds into wind farms yielding 1 percent per year when it is borrowing money in international financial markets at 5 or 10 percent. [...] The discount rate should be primarily determined by the opportunity cost of capital, which is determined by the rate of return on alternative investments. (Nordhaus 2013a, 187)

To grasp the basic idea behind this way of thinking about discounting, consider a government climate change project that costs $100 today and prevents $1,000 in climate damages in 100 years. Nordhaus’ way of judging whether this project is “economically desirable” (76) is to check whether the government could have delivered more than $1,000 in benefits in 100 years by investing the $100 in some alternative investment. If all other forms of investment are already effectively competing against one another and thus have the same marginal rate of return (something Nordhaus assumes), then a neat way to check mitigation against other investments is to discount the benefits offered by mitigation using a discount rate equal to the rate of return offered in capital markets. Suppose, for example, the interest rate available in capital markets is 6% per year. Then we can multiply the $1,000 future benefit by the corresponding discount factor, \( \left( \frac{1}{1.06^{100}} \right) \), and subtract the $100 present cost of producing that $1,000 benefit. If the result of this exercise is a positive number, that indicates the climate change project is the better of the two. If the result is a negative number—which in this case it is—then the government would do better to make a conventional capital investment instead. Indeed, the government could either invest the full $100 to yield roughly $40,000 worth of benefits in 100 years, or it can invest just $2.50 today in order deliver $1,000 in 100 years. Either way, Nordhaus claims that countries should investigate whether a climatic investment beats a conventional financial investment, and the way to determine this is by discounting the benefits of the former using the rate of return offered by the latter. Countries should then engage in the most lucrative climate projects until the remaining projects have discounted benefits equal to or less
than the project’s cost. Once they have done that, they have achieved the ‘optimal’ level of abatement and the global temperature increases associated with this level of abatement is the ‘optimal’ increase.

In order to implement this method of climate policy analysis, one needs a way to predict the relevant capital market interest rates far into the future. For this, Nordhaus enlists an economic model known as the Ramsey-Koopmans-Cass model. The standard macroeconomic approach to Ramsey analysis (as I will call it) is to suppose that the economy can be thought of as driven by the behavior of a single infinitely-lived representative agent who must choose between using income to fund consumption today or instead investing it to fund consumption tomorrow. It is supposed that there is a set of feasible intertemporal consumption streams of the form \( \mathcal{C} = \{c_t^n, c_{t+1}^n, c_{t+2}^n, \ldots\} \) over which the agent can choose, and that she adjusts her consumption and investment decisions so as to place herself on her top-ranked stream. (Here, each \( c_t^n \) represents the agent’s level of consumption at time \( t \) along stream \( n \mathcal{C} \).) It is further supposed that the agent’s preferences over consumption streams can be represented by a discounted utilitarian value function of the form:

\[
V_t = \sum_{t=0}^{\infty} u(c_t) \cdot \left( \frac{1}{(1 + \delta)^t} \right) \tag{1}
\]

This value function ranks consumption streams by using its \( u \)-function to represent the agent’s consumption-derived utility at each time, discounting utility that occurs in the future, and then summing these utility numbers across times. The representative agent is said to be impatient at a given time if her utility discount rate, \( \delta \), is greater than zero at that time (and thus if her corresponding utility discount factor, \( \left( \frac{1}{(1+\delta)} \right) \), is less than 1). The utility function \( u \) is typically assumed to take the isoelastic form:

\[
u = \frac{c_t^{1-\eta}}{1-\eta} \tag{2}
\]

In this utility function, \( \eta \) is a measure of the function’s curvature, which reflects both the degree to which the agent is averse to risk to her consumption levels and the degree to which she is averse to fluctuations
in her consumption across time. When $\eta = 0$, the representative agent is neutral with respect to risk and also indifferent to consumption inequalities across the different times in her life. (Risk- and inequality-aversion are conceptually distinguishable from impatience, and from each other, but the standard utility function in climate economics nevertheless forges an ironclad link between risk and inequality aversion (Kelleher and Wagner 2018).) As $\eta$ approaches infinity, the agent is increasingly averse to consumption fluctuations and increasingly prefers to maximize the lowest consumption level across all times she is alive.

After calibrating these two preference parameters to observed behavior in the real world, macroeconomists use their models to predict real-world future interest rates. They do this via a famous equation known as the Ramsey equation. Actually, there are two Ramsey equations. Here is the first, which can be derived from equation (1):

$$\rho_t = \delta + \eta \cdot g_t$$  \hspace{1cm} (3)

Call this equation the \textit{Ramsey formula}. Here is how it works. Suppose for the sake of argument that $\mathcal{C}$ is the business as usual consumption stream. For each time period $t$ along stream $\mathcal{C}$, there will be a rate of growth between the agent’s consumption at $t$ and her consumption in the very next time period, $t + 1$. Designate this rate of growth between $c_t$ and $c_{t+1}$ as $g_t$. Now by plugging $g_t$, $\delta$, and $\eta$ into equation (3), one generates period $t$’s consumption discount rate, $\rho_t$. A consumption discount rate is a sort of ‘hurdle rate’. To illustrate, suppose the consumption discount rate for period $t$ along the business as usual consumption stream $\mathcal{C}$ is 0.06, or 6%. This means that if it is possible for the representative agent to decrease her period $t$ consumption by one dollar and receive a greater than 6% return on that investment in period $t + 1$, then the agent’s preferences are better satisfied by investing the dollar instead of consuming it. In other words, if the prevailing rate of return on investment ‘hurdles’ over the agent’s consumption discount rate, then the agent’s preferences are better satisfied by consuming slightly less now in order to invest and consume a bit more later.

By calibrating their representative agent’s discounted utilitarian preferences so that they approximate the preferences of the average real-world agent, macroeconomists observe as their model’s agents
move from feasible consumption stream to feasible consumption stream until they settle on a stream along which their consumption discount rates in every time period equal those respective time periods’ available rates of return on investment. This gives the second Ramsey equation:

\[ r_t = \delta + \eta \cdot g_t \quad (4) \]

This equation, which is sometimes called the Ramsey rule, says that agents with discounted utilitarian preferences will make savings and consumption decisions so as to ensure that each time period’s consumption discount rate (as given by the righthand side of (4)) equals that time period’s rate of return on conventional capital investment (as given by the lefthand side of (4)). The reasoning is intuitive: if equilibrium consumption discount rates did not equal equilibrium rates of return in all time periods, the representative agent could move to a more preferred consumption stream by adjusting her consumption or saving behavior in at least one time period. The Ramsey rule is an ‘optimality condition’ in the sense that if one has discounted utilitarian preferences, then a necessary condition for residing on one’s most preferred consumption stream is that the Ramsey rule equation actually hold true.

Nordhaus’ approach to projecting real interest rates is actually a bit different from the standard approach I have just described. Instead of assuming a representative agent, he assumes a representative society. He knows that real-world decisions about which consumption streams to settle on are profoundly shaped by government policies (e.g. tax policy), and so instead of modeling the planning behavior of an agent (or, equivalently, an identical set of agents), Nordhaus models the decision-making of societies by using the following social discounted utilitarian value function:

\[ V = \sum_{t=0}^{\infty} N_t u(c_t) \cdot \left( \frac{1}{(1 + \delta)^t} \right) \quad (5) \]

\[ ^4 \text{It is unfortunately not common in the climate economics literature to distinguish between what I have called the Ramsey formula and the Ramsey rule. But they are importantly distinct propositions and should be kept distinct. For the rare paper that does so, see Dasgupta (1982, 279, 284).} \]
With the addition of the population variable $N_t$, this formulation moves beyond the representative agent conceit by explicitly acknowledging that multiple agents exist at each time. In Nordhaus' model, the utility discount rate, $\delta$, is now the rate at which society discounts the utility (which is still given by an isoelastic $u$-function) of future people. Despite these differences, the Ramsey formula (equation (3)) and the Ramsey rule (equation (4)) carry over from the representative agent framework to the one employing a social value function.\(^5\) This is related to one of Nordhaus’ key assumptions: he assumes that real-world societies have used public policy tools to bring it about that the equality expressed by (4) holds true in the real world. From this assumption, he proposes to deduce the values of $\eta$ and $\delta$ that characterize these societies’ policy preferences. Of course, there is an infinite number of pairs of $(\eta, \delta)$ that could equally well solve a Ramsey rule equation with an observable business-as-usual real interest rate on the lefthand side and an observable per capita consumption growth rate on the righthand side, and Nordhaus has modified his selections over the years (see for example Nordhaus 2008, 50). Still, his overall aspiration has remained the same:

We note first, as discussed earlier, that the interpretation of the economic parameters is that they are designed to provide the most accurate projections rather than to be normative in nature. (Nordhaus 2008, 50)

Here, then, is how Nordhaus' models use a social value function to simulate perfect competitive markets for greenhouse gas abatement. First, he seeks to infer values for $\delta$ and $\eta$ by assuming that societies optimize their conventional investments. The parameter values so inferred are descriptive: they describe aspects of the social policy preferences that steer the real-world macroeconomy. He then inserts these parameters into his IAM, which is independently programmed to identify the full set of feasible future consumption streams. Nordhaus then asks his model to determine which feasible consumption stream maximizes the descriptive social value function subject to the constraint that countries do not invest in greenhouse gas mitigation. This gives the

\(^5\) In the context of a social discounted utilitarian value function, the Ramsey formula expresses the hurdle rate for an increment in total consumption, rather than an increment in the representative agent’s consumption, and $c_i$ now represents average per capita consumption.
business-as-usual consumption stream. Nordhaus’ model then computes this stream’s time-path of consumption discount rates, and finally the Ramsey rule is invoked once again to equate these discount rates with the business-as-usual time-path of rates of return on capital investment. Countries can now discount prospective future benefits of greenhouse gas mitigation with these interest rates in order to test climate investment projects against alternative investment opportunities elsewhere in the economy. If we assume that marginal mitigation costs in the model are given by the costs as they would be if countries were willing and able to jointly invest in meaningful greenhouse gas abatement, then by supposing that countries invest in mitigation until the discounted benefits of marginal mitigation equal present marginal abatement costs, Nordhaus’ model identifies the additional mitigation that the world’s countries would collectively undertake in the absence of the three barriers that currently ‘triply persuade’ them to underinvest in climate change mitigation.

Thus for Nordhaus, ‘optimal’ climate policy is the policy that each of the world’s countries would itself support, given its actual current policy preferences, if all the countries could work together to nullify the impulse to free-ride on the efforts of others. Importantly, Nordhaus admits that he is only projecting what countries would do under more favorable circumstances, rather than opining on what countries should morally do. For example, in criticizing the ethically-informed approach to discounting and climate policy analysis adopted by Nicholas Stern’s British government-commissioned *The Economics of Climate: The Stern Review* (Stern 2007), Nordhaus (2007, 692) says, “[I] find the ethical reasoning on discount rates in the Review largely irrelevant for the actual investments and negotiations about climate change”.

But then since Nordhaus studiously avoids ethical analysis, it is very unclear why he feels entitled to assign the label “optimal” to the outcome of his quasi-predictive exercise and to associate that outcome with “idealized” and “economically desirable” policy (Nordhaus 2008, 14; Nordhaus 2013a, 76). At times, Nordhaus has acknowledged this problem for his frequent policy-prescriptive declarations. In a paper that has been cited much less frequently than his much more famous review of the *Stern Review*, Nordhaus acknowledges that his model is simply “an algorithm for finding the outcome of efficient competitive markets” and that unless more robust ethical assumptions are incorporated into it—something Nordhaus himself has always refused to do—then the
model’s “optimal” outcome is actually “purely algorithmic and has no compelling normative properties” (Nordhaus 2013b, 1111).

II. “IF I COULD REWIND THE CLOCK TWENTY YEARS, I WOULD HAVE WRITTEN THIS DIFFERENTLY.”

In 2012 Nordhaus was a commentator at John Broome’s Tanner Lectures on Human Values at the University of Michigan. Broome is a former-economist-turned-leading-moral-philosopher, and he has been one of philosophy’s staunchest critics of Nordhaus’ approach to climate economics. At one point, Nordhaus stood in front a whiteboard on which Broome had previously drawn a diagram very similar to figure 1. Nordhaus then said:

Now, so then we have the business-as-usual point [point BAU], and the idea here being that that’s inside [...] the frontier because of inefficiencies—in this particular case climate externalities. Then we have what I think of as the Pareto region here. [...] Those are in principle places where you could improve the consumption of both future and present generations. [...] So, John [Broome] is correct to say that the policies that were emphasized in the Stern Review and in my work were ones that were up here to the northwest of the Pareto region. [...] Now in my own work, if I could rewind the clock twenty years, I would have written this differently—I would have emphasized differently, which is [to say] that there are two kinds of things going on here: one is going to the frontier [starting from point BAU], which you can think of as climate policy; and then where you are on the frontier, which is distributional policy. (Broome and Nordhaus 2012, 37:10–43:00 min.)

Broome drew his diagram in part to illustrate what happens when emitters are permitted to emit without any regard for the external costs they create. One thing that happens is that emitters’ greenhouse gas emissions lead to Pareto inefficiency. When an economist says that a situation is Pareto efficient, it means it is not possible to make any person better off (as compared to that situation) without thereby worsening someone else. By that definition, every point on figure 1’s consumption possibility frontier is Pareto efficient: for every such point, P, if the economy were located at P, any move away from P would involve making at least some party worse off than they were at P. Conversely, a point is Pareto inefficient when it lies below the

---

6 Economists use “Pareto optimal” and “Pareto efficient” synonymously.
consumption possibility frontier; in that case, it is possible to move to the frontier in a way that improves at least one person without making anyone else worse off. Call such a move a *Pareto improvement*.

Now consider figure 1’s point BAU. Note first that any point on or under the consumption possibility frontier represents a consumption stream with just two time periods—*present* and *future*. Being under the frontier, the consumption stream represented by BAU is Pareto inefficient: it is possible to move to a new consumption stream that makes at least one time-period’s people better off without making anyone else worse off—a Pareto improvement. The full set of consumption streams offering Pareto improvements falls within what Nordhaus termed the Pareto region. When greenhouse gasses are completely unregulated and untaxed, those who suffer the bad effects of greenhouse gas emissions might well be willing to pay much more for

**Figure 1. Illustrating Intergenerational Consumption Possibilities**
meaningful abatement than the emitters would be willing to accept by way of compensation to reduce their emissions. This because the emitters will emit right up until their declining marginal private benefit of the emitting activity equals the marginal private costs of that activity, at which point those who suffer the external costs of the activity may be experiencing so much pollution that they are willing to pay a lot to prevent some of it. And that is why Broome drew the BAU point as lying under the consumption possibility frontier.

Suppose BAU represents our current economy’s status quo consumption stream. In that case, there are Pareto-improving trades that could make both present and future people better off. Such trades involve exchanging resources across time: future generations would be willing to forego some material wealth in exchange for an improved climate. If they were possible, a series of such intertemporal trades would shift the economy from BAU to a point on the consumption possibly frontier that is within the Pareto region. And indeed Nordhaus describes just such a policy in his review of the Stern Review: we can take out a loan today to compensate ourselves for undertaking climatic investments, and the loan can be kicked down the road until future generations arrive to pay it off (Nordhaus 2007, 695).

What I have just laid out is the background behind Broome’s claim “that the problem of climate change can be solved without anyone making a sacrifice” (Broome 2010, 102). Broome conceives of the “problem of climate change” as consisting in what economists call a Pareto-relevant externality, or an externality that keeps the economy operating underneath the consumption possibility frontier, which leaves all people worse off than they could otherwise be (Buchanan and Stubblebine 1962; Baumol and Oates 1988, 16). On this understanding, the problem of climate change consists in the fact that there exist paths of emissions abatement that, when coupled with a corresponding intertemporal transfer of resources, could move the economy onto the frontier. Once on the frontier, any further problem concerns whether or not “it is better to have more resources in the hands of the poor and of future generations, rather than in the hands of the current rich”. And “[t]his matter of distribution has little to do with climate change”, according to Broome (2012, 46).

Thus in at least one sense, Broome and Nordhaus agree that (as Nordhaus puts it) climate policy concerns “going to the frontier” whereas distributive policy concerns the evaluation of “where you are on
the frontier” once the economy gets there (Broome and Nordhaus 2012, 39:15 min.). They also seem to agree that it is strongly worth considering routes to the frontier that could win political support in the world as it is. The key difference, it seems to me, is that while Broome is trying to convince present people to undertake a climate policy that will make them no worse off, Nordhaus is trying to convince governments to undertake a climate policy that comports with their revealed intertemporal preferences. As Nordhaus puts it when justifying his approach to climate economics in his review of the Stern Review, “When countries weigh their self-interest in international bargains about emissions reductions and burden sharing, they will look at the actual gains from bargains, and the returns on these relative to other investments” (Nordhaus 2007, 692). Yet the fact remains that if one wishes to separate climate policy and distributional policy in the way Nordhaus claims he does, and if one wishes to restrict oneself to climate policy only, then one cannot insist that any given policy-induced route to the consumption possibility frontier is more ‘optimal’ than another.

Even if Nordhaus is wrong to label his particular route to the Pareto frontier as ‘optimal’ a key redeeming point is that Nordhaus’ non-normative approach to discounting is precisely what Broome needs to help identify the Pareto-relevant climate externality in the first place. This is because monetized climate damages can be interpreted as future people’s willingness to pay to prevent those damages. (If one stands to lose $100 from climate change along a business as usual climate-economy pathway, then one should be willing to pay up to $100 to move to a new pathway that avoids those damages.) By discounting future marginal damages using the real interest rates that would result along the business as usual pathway, one converts future willingness to pay to prevent those damages into present-dollar terms. And it is the present value of future people’s willingness to pay that must be compared with present people’s willingness to accept compensation if one wishes to illuminate the Pareto-relevant climate externality.

III. CONCLUSION
Neither Broome nor Stern believes that one must renounce one’s moral compass in order to flag the important fact that much greenhouse gas mitigation can be Pareto-improving (see Stern 2015, 202–205). Yet it is precisely Nordhaus’ descriptive approach to discounting—or perhaps some alternative but still descriptive approach—that is required to
reveal the full extent of the Pareto-relevant climate externality. It therefore seems possible to reconcile Nordhaus' broad approach to discounting with the alternative ethical approach championed by Stern and allied moral philosophers. That alternative ethical approach aims to identify the point on the consumption possibility frontier that is (in some sense requiring further elaboration) morally optimal. It seems to me that each of these approaches to climate economics can shed light on a policy-relevant aspect of the overall climate change problem. But if this is right, it is also true that Nordhaus’ ‘optimal’ policy largely reflects one person’s sense of which aspects should be emphasized in real-world policy discussions, rather than an impartial finding from a prize-winning economic model.

REFERENCES


This reconciliation resembles to some degree the one proposed in Howarth (2003) and perhaps also in Posner and Weisbach (2010). But neither puts things as I have, and I think there remain differences of interpretation between us. For an analysis more along the lines I've given here, and to which I am indebted, see Goulder and Williams (2012).


**J Paul Kelleher** is Associate Professor of Bioethics and Philosophy at the University of Wisconsin-Madison. His research focuses on ethical and philosophical dimensions of public policy, especially health policy and climate policy. He is currently writing a book on climate change and intertemporal welfare economics.

Contact e-mail: <paul.kelleher@wisc.edu>

LACHEZAR GRUDEV
University of Freiburg

Friedrich A. Hayek is often considered to have been an economist and a social philosopher who was a strong free-market proponent and who argued that markets are perfectly efficient in generating an optimal allocation of resources. In his recently published book, F. A. Hayek: Economics, Political Economy and Social Philosophy, Professor Peter J. Boettke argues that such claims represent a common misconception about Hayek’s central ideas. Boettke’s book aims to clarify these misconceptions because, he argues, they prevent us from understanding Hayek’s arguments. For Boettke, these arguments are still relevant to contemporary discussions and can contribute a lot to our debates, particularly in the methodology of the social sciences and in technical economics.

This review will discuss those points in Boettke’s book that we as economists lack in our model-based thinking, but that are, nevertheless, highly relevant to our efforts to grasp reality. The book is highly recommended for academics concerned with technical economics as it is an invitation to talk about thinkers, whose theories have either been misinterpreted or unjustly remained forgotten in the old books.

Boettke’s book conveys the simple message that the history of economics matters since it provides valuable answers to highly topical questions, such as the role of the institutional framework on the economic process. Boettke argues that Hayek was involved in many debates during the 1930-1960 in economics and political economy which gave rise to a renewed focus on the institutional framework within which economic activity takes place. After this period Hayek explored the discovery and learning aspects of alternative institutional arrangements, which Boettke defines as epistemic institutionalism. Here, the fundamental question concerns the institutional prerequisites for learning and error correction among individuals in a society. By focusing on this, the book derives important implications for the methodology of...
the social sciences, for analytical economics, and for practical public policy.

**Hayek as a Technical Economist: The Coordination Problem**

Fundamental to understanding Hayek’s epistemic institutionalism are the chapters “Hayek: An Overview of His Life and Work” and “The Anatomy of an Economic Crisis: Money, Prices, and Economic Order”. Here, Boettke points out that Hayek was influenced by the earlier generations of the Austrian School, particularly Friedrich von Wieser and Ludwig von Mises. In his early period, Hayek incorporated Austrian concepts into his explanation of the exchange process and the structure of production. At the same time, Boettke delineates the main difference between Keynes and Hayek. For Hayek, macroeconomics obscures the economic problem which should be viewed as a coordination process of economic activity through time. Relative prices guide future exchange and production decisions, and thus agents’ economic plans, because producers utilize the price system with the intention to “make rational economic calculation about alternative courses of action in commercial activity” (p. 42).

In his technical books on cyclical fluctuations, Hayek showed the fatal consequences that might result when relative prices fail to fulfill this information function. Based on Böhm-Bawerk’s concept of a production period and on Mises’s monetary theory, Hayek pointed out that if the money interest rate is manipulated and artificially lowered below the natural interest rate, then producers would concentrate their resources on ventures that had not been profitable before the change in the interest rate. This affects relative prices which, in turn, influences the decision-making of the producers and alters the production structure of the economy. The new production structure is not sustainable because commercial banks would at some point be compelled, first, to raise the interest rate due to liquidity reasons and, second, to curtail the credit provision for new ventures to the producers. Market participants would realize that they had been misinformed. The result would be a crisis and a depression that Hayek characterized as necessary adjustment processes to the pre-boom production structure of the economy.
HAYEK AS A POLITICAL ECONOMIST: THE INSTITUTIONAL INFRASTRUCTURE

During the following decades, Hayek refined his ideas on the coordination process in the economy. This is illustrated most convincingly in Boettke’s chapter “Hayek on Market Theory and the Price System”. This chapter focuses on how Hayek integrated the role of institutions regarding the acquisition and transmission of knowledge into his analysis of the coordination of the economic agents’ plans. What is fundamental for Hayek is that knowledge is never given to any single mind, but that it is dispersed throughout society. This allows each member of a society to possess a small bit of an incomplete and subjective knowledge pertaining to a specific time and place. Boettke points out that for Hayek the price system represented a fundamental “mechanism for communicating information” about the relative scarcity of resources, and enabled individuals to act “as if they possessed more knowledge than a single mind could grasp without the use of the price system” (p. 86).

Boettke underlines that for Hayek knowledge incorporated in prices was not sufficient for the spontaneous coordination of economic plans. The subjective knowledge about objective facts, such as the alternative uses of one good or its different substitutes known to the acting agent, is acquired not only from the price system. Hayek considered the competitive market process to be the central institution for knowledge acquisition. The reason for this is that competition with its trial and error experimentation works as a discovery procedure. This procedure reveals facts, such as new ways of production, of which the individual has so far been unaware. Boettke claims that for Hayek competition is “an institutionally-contingent social process” (87) because, in order to work as a discovery procedure fostering innovations and experimentations, specific institutional prerequisites must be fulfilled: property rights, a functioning price system, a limited government, and the rule of law.

Boettke uses this discussion to clarify specific misconceptions in modern informational economics. This research program claims that Hayek treated knowledge of prices as the only relevant type of knowledge and that this knowledge is sufficient for attaining equilibrium. Boettke traces this claim to explicit sections in the papers of Grossman (1976), and Grossman and Stiglitz (1976, 1980), who argued that economic agents, who know the price system and act out of
self-interest, can generate resource allocations in a given economy. But this economy should not be viewed as automatically stable “because prices are revealing so much information that incentives for the collection of information are removed. [...] It is not enough for traders to observe only prices” (Grossman 1976, 585) (105). Boettke shows that this stands in complete contrast to Hayek’s statement that the price system is not sufficient for achieving equilibrium. Individuals have to possess additional knowledge in order to adapt their plans to the changing circumstances.

In the chapter “Hayek and Market Socialism”, Boettke asserts that these arguments provide the basis for Hayek’s response to the intellectual challenges facing the advocates of socialism, particularly Oskar Lange and Abba Lerner. They aimed at developing a “market socialism”, in which a figure akin to the Walrasian auctioneer would emulate “the tâtonnement process under general competitive equilibrium” (p. 23). Here producer markets can be abolished because their vulnerability to cyclical fluctuations gives rise to depressions, and thus to unemployment and poverty. The prices for producer goods can be computed based on the marginal theory of value. A ‘central planning board’ can be established that replaces the unstable markets and that communicates to the managers of state-owned firms how to produce in an optimal manner.

Boettke asserts that Hayek treated such planning processes as impossible because price imputation is not a trivial procedure. Producers rely on many signals that guide them in deciding which methods to use for production. This requires market participants to be adaptable to changing circumstances; it requires an adaptability, which is based on subjective knowledge regarding specific methods of production. Price competition creates an incentive for entrepreneurs to acquire knowledge regarding the best and cheapest methods of production. Hence, the average cost curve is not a given fact for entrepreneurs, as the market socialists think, but a result of the complicated discovery procedure taking place within the market process. While price competition fosters an incentive to acquire knowledge about the best method of production, for Hayek, institutions enable entrepreneurs to acquire this knowledge pertaining to the specific context in which entrepreneurs evaluate their past decisions and make new decisions, with the final goal of coordinating their plans with those of the other market participants.
HAYEK AS A SOCIAL PHILOSOPHER: THE LIBERAL ORDER

Let us conclude with the main message of the chapters “The Reconstruction of the Liberal Project” and “The Hayekian Legacy”. These chapters point out why Hayek’s intellectual legacy is topical today when populist movements from the left and the right with their nationalistic and anti-globalist rhetoric are challenging the foundations of free societies. These two chapters can be read as an appeal particularly to economists as social scientists. An appeal to put aside formalism and rational-choice models so as to understand Hayek’s main message: that the problem of economics is the coordination of dispersed knowledge in society. Coordination requires institutions that enable people to learn what they want to learn, so that coordination of plans which are fundamental to prosperity and welfare can be promoted. This explains why institutions play an important epistemic role in encouraging peaceful cooperation among different individuals independently of their language, ethnicity, race, religion and geography. The theorists of liberalism have been clarifying how this institutional framework creates the conditions for such a peaceful coordination and increases the costs of discrimination based on religion, gender, and race. Boettke argues that a basic idea of liberalism has been the normative vision of the classical political economists since Adam Smith.

This book is highly recommended not only for Hayek experts and academics interested in Hayek’s intellectual legacy, but also, and particularly, for economists engaged in technical economics. Boettke’s book advances Hayek’s epistemic institutionalism as an important constructive program, because it accentuates the relevance of institutions to the coordination of market participants, and thus to how institutions shape economic processes. By clarifying the common misconceptions about Hayek, Boettke’s book invites us to reread Hayek, putting our ideological glasses aside. Boettke convincingly describes Hayek’s story as that of an economist and a social philosopher who was also concerned with understanding the reality by seeking an answer to a very simple question: what enables us, fallible, but capable individuals, to coordinate our “affairs with one another and realize productive specialization and peaceful cooperation” (229), giving rise to the tremendous increases in wealth and prosperity observed in the last two centuries. Understanding Hayek’s answer to this question and overcoming our prejudices about Hayek’s theoretical approach will contribute to the progress of economic science.
REFERENCES

Lachezar Grudev is a PhD student in economics at the University of Freiburg, Germany. His research interests include history of economic thought, history of business cycle theories, economic methodology, constitutional economics and public choice. He is a member of History of Economics Society and Wilhelm-Röpke-Institute. He is also a Mercatus Center Adam Smith fellow during the 2018-2019 academic year.
Contact e-mail: <lachezar.grudev@vwl.uni-freiburg.de>

JOSEPH HEATH  
*University of Toronto*

The past decade has been exceptionally kind to libertarians. Once confined to an isolated and oppressed intellectual minority, thumbing through dog-eared copies of *Anarchy, State, and Utopia*, while muttering darkly about ‘forced labor’ and ‘self-ownership’, the view has since undergone something of a renaissance. Apart from having developed both ‘left’ and ‘right’ variants, the theory has largely escaped from the shadow cast by Robert Nozick, so that theorists are now arguing for libertarian conclusions from a wide range of different premises.\(^1\) The internet has played a crucial role in reinvigorating the movement. Libertarians have formed a number of thriving online subcultures, even attracting many readers and commentators with little sympathy for the view, but intrigued by the vibrancy and intellectual challenge of the positions being discussed. Central to the latter phenomenon has been the economics blog *Marginal Revolution*, which has served as an extraordinary resource for this community. And central to the blog has been George Mason University professor Tyler Cowen who, as a self-described “infovore” (2010), not only generates a phenomenal volume of provocative posts, but also sifts through other writing at an almost superhuman rate, linking, quoting, and recommending literature on a surprisingly broad range of topics.

But above all, *Marginal Revolution* has become an invaluable resource for a diffuse group of like-minded individuals, who subscribe to a political position that I like to think of as ‘internet libertarianism’—not because it has anything to do with the internet, but just because, if you come across someone on the internet claiming to be a libertarian, they will almost always hold something like this position. It is less of a worked-out theory than it is a collection of beliefs. Central to it is the view that government is, in general, bad, while markets are, almost

\(^1\) On left libertarianism, see Vallentyne and Steiner (2000), also Otsuka (2003). For doctrinal diversity, see Tomasi (2012), or Levy (2015).
always, good. This normative judgment is taken to follow from two claims, which are seldom clearly distinguished from one another. First, there is the view that markets are good because they promote welfare, while the state is bad because it is inefficient. The second is the view that markets are good because they promote liberty, while the state is bad because it is coercive.

It is this second claim, of course, that confers the title ‘libertarian’ upon the position, and that distinguishes it from mere liberalism. After all, ‘liberals’ (in the political-philosophy sense) have for a long time touted the welfare-enhancing qualities of markets. But they have also, for a long time, made the case for the growth of the state—whether it be through regulation, the development of the social safety net, or the provision of public and club goods—in terms of the welfare-enhancing consequences of these interventions. This makes it very difficult to motivate much anti-government rhetoric just by insisting on the importance of efficiency. As a result, those who want to set themselves apart as offering a libertarian defence of the market must say something that goes beyond the standard ‘invisible hand’ argument.

One way of toughening up the stance in favor of liberty is to jettison the commitment to welfarism. This is the old-school strategy, which one can find in writers like Ayn Rand and Nozick (both of whom regarded the fact that markets make people happy as somewhat incidental to their concerns, and irrelevant from the moral point of view). But this is too harsh for the modern ‘bleeding heart’ libertarian, who cares about liberty, but also wants people to be happy. These two positions are, unfortunately, extremely difficult to reconcile. Particularly if the commitment to happiness is interpreted in consequentialist (e.g. utilitarian or quasi-utilitarian) terms, it tends to instrumentalize the commitment to liberty, so that liberty winds up being good if it promotes good consequences, such as general happiness, but dispensable in cases where it does not. This basically rules out extremism in defence of liberty, which, in turn, takes all the fun out of being a libertarian.

Because of this, much of contemporary libertarianism winds up being a paper tiger. It typically starts out with a lot of big talk about contract, freedom, and the blood of patriots, but when push comes to

---

2 See “Man’s Rights” in Rand (1964), also Nozick (1974).
shove, social welfare winds up doing all the normative work.\(^3\) The commitment to liberty often shows up as a downstream consequence of some empirically dubious claims about the ineptitude of public servants, and the hidden virtues of private ownership. Central to this reconciliation is the tacit denial that individuals can get locked into states of ‘dysphoric freedom’, on the model of the prisoner's dilemma, where free choice generates outcomes that are, by common admission, making everyone unhappy.

Cowen’s recent book, *Stubborn Attachments*, is not immune to this difficulty. This short treatise represents Cowen’s attempt to take a step back from the barrage of information that he processes every day, in order to provide a more general statement of the philosophical commitments that inform his thinking. For our purposes, it can be thought of as a type of internet-libertarian manifesto. The central question is, what does it bring to the table? The answer is that, while it contains a (perhaps surprisingly large) number of conventional ideas, there is one unconventional one, which Cowen uses in an attempt to resolve the fundamental internet-libertarian dilemma. To preview: Cowen is opposed to temporal discounting. As a result, he wants to ‘expand the circle’ of moral concern to include all future people. Doing so offers a rather unexpected strategy for resolving the tension between his twin commitments to both consequentialism and a deontological conception of individual rights.

Before getting into the details, a warning: professional philosophers will no doubt find Cowen’s book exasperating. Two major problems stand out. First, he is not worried at all about the foundations of his normative commitments, he simply posits them. Thus he posits a commitment to maximizing welfare, along with a host of other goods, and he posits a commitment to “human rights” (2018, 22), but does not really say where they come from. From a philosophical perspective, he might as well be pulling these things out of a hat. Second, many of the commitments that he posits seem incompatible, but instead of worrying about this, he tends to engage in hand-waving of one sort or another. Most obviously, he posits a set of “absolute” (2019) rights (or “(nearly) absolute” (p. 22) in some versions), but then declares himself to be a pluralist about value. Pluralism, however, implies a willingness to make tradeoffs, which in turn makes it difficult to see how the commitment to

---

\(^3\) A good example of this is Easterbrook and Fischel (1991).
rights could be absolute. Furthermore, he does not enumerate the rights that he is committed to—the only specific one he mentions is the right of innocent babies not to be murdered. Without knowing how he feels about more controversial rights, such as the right to property—which he says nothing about—it is impossible to know how much conflict there is likely to be between these rights and the other political values he espouses.

But of course, the book is not intended to be an academic treatise. It is an attempt to organize and provide some structure to a set of widely held ideas, and to provide for his many readers a better sense of ‘where he’s coming from’. It is therefore unfair to evaluate it by the persnickety standards of academic philosophy. One might treat it then as a profession de foi, and not worry too much about how it hangs together. This is what many commentators have done, and I have little to add to this discussion. What I propose to do, instead, is to offer something of a ‘rational reconstruction’ of the central argument, presenting it in a more systematic (and defensible) form than Cowen does. I do so because there is an interesting philosophical idea to be found there, which merits serious consideration.

The first thing to note is that, despite going to some lengths to distance himself from both utilitarianism and preference-based welfare consequentialism, Cowen’s fundamental moral commitments are consequentialist. One can see this in his discussion of economic growth, which occupies the first two chapters of the book. He adopts from Frank Knight the image of a Crusonia plant, “a mythical, automatically growing crop, which generates more output each period” (p. 28). Other plants may produce greater yields in the short term, but eventually they wither and die. The Crusonia plant just keeps producing, increasing its yield at a steady rate of (say) 5%. Given a choice, Cowen observes, between tending one or another crop, one should focus on the Crusonia plant, because in the fullness of time it is guaranteed to outperform any other. More generally, it offers us the opportunity to resolve what he refers to as “aggregation problems” (p. 29). What he has in mind are the distributive conflicts that can be generated through straightforward application of the utilitarian calculus, where the interests of some are

---

4 The Crusonia plant bears a striking resemblance to the shmoo, which has figured in certain philosophical discussions.
sacrificed in order to produce benefits for others.\(^5\) Cowen does not propose any formula for resolving these conflicts, instead he urges us to focus on avoiding them, by putting as much of our energy as we can into the cultivation of Crusonia plants. While this may generate conflict and resistance in the short term, if we are sufficiently stubborn in our attachment to them, we can rest assured that, in the long run, they will generate Pareto improvements.

The Crusonia plant is, of course, not entirely fictitious, it is a thinly veiled metaphor for economic growth (p. 29). As a result, Cowen's 'theory of justice', stated as simply as possible, is that we should stop fighting over who gets what, and focus on arranging things so that everyone can get more. This is, of course, a view quite widely held by economists—indeed, it is so conventional that I was somewhat surprised to hear it coming from Cowen. Whatever its merits, however, it is important to note that it is very far from being libertarian. Indeed, from a strictly normative point of view, it more closely resembles Marxism. Marx, it may be recalled, was also quite pessimistic about the possibility of any moral resolution to the distributive conflicts that arise from the joint character of economic activity. He thought that these conflicts were settled in the streets, not at the level of philosophical theory. This is why the history of all hitherto existing society is the history of class conflict (and why morality is nothing but ideology). Capitalism, however, promises to put an end to all this, by producing goods in such abundance that it renders distributive conflict otiose, and therefore permits the formation of a classless society. Marx, of course, believed that this abundance would allow society to abolish the market, whereas Cowen sees quite rightly that it will not. But apart from these institutional details their views have essentially the same normative structure.

The discussion of Crusonia plants also brings to the fore the consequentialism at the heart of Cowen's view. One does not need a growing economy to satisfy the basic deontological constraints that a libertarian like Nozick valued the market for. The central attraction of growth is that it produces increased preference-satisfaction. Cowen is careful to point out that increased consumer satisfaction is not the only value. Non-market values that are worth pursuing include increased

\(^5\) He defines ‘aggregation’, tendentiously, as “how we resolve disagreements and how we decide that the wishes of one individual should take precedence over the wishes of another” (p. 19).
leisure, health, life expectancy, political freedom, and environmental quality. But as he points out, all of these are positively correlated with economic growth, and so expanding the list of goods beyond the merely economic (to what he calls “wealth plus” [p. 30]), does not require any important adjustment of strategy. By maximizing growth, we maximize the sum of good things:

The bottom line is this: the more rapidly growing economy will, at some point, bring about much higher levels of human well-being—and other plural values—on a consistent basis. If some set of choices or policies gives us a higher rate of economic growth, those same choices or policies are akin to a Crusonia plant (p. 41).

Again, this is all fine, but it is also a fairly common liberal sentiment. None of it leads to a particularly libertarian view. Indeed, this framework is one that permits, and perhaps even recommends, any cost-benefit justified state intervention in the economy. This puts us in Cass Sunstein territory. So how do we get from this to the distinctive concern over liberty? One approach would be to build freedom into the consequences, in the way that Amartya Sen does (“it’s not the functionings that matter, dummy, it’s the space of possible functionings”). Cowen rejects this, on the grounds that it loses the deontological flavour of our commitments to liberty and other individual rights. But what resources does Cowen have, other than standard rule-utilitarianism?

It is at this point that the one genuine peculiarity of Cowen's view surfaces and begins to play an explicit role in the argument. This is his view of temporal discounting. Those who follow Cowen's academic work will know that he published a pair of articles in the early 1990's—one co-authored with Derek Parfit—in which he defended the view that it is morally impermissible to treat the time at which a particular good consequence occurs as a morally salient feature of that event (Cowen and Parfit 1992; Cowen 1992). This commitment to temporal neutrality is a widely held view among philosophers, but rather uncommon among economists. The reason is that economists are accustomed to thinking about investment decisions (as well as economic growth), and worry about the consequences of introducing infinite utility streams into the calculus of costs and benefits. After all, without some sort of discounting, it would be possible to justify arbitrarily large one-time costs, in order to secure a stream of arbitrarily small but recurring
benefits. Thus economists have habitually bounded the utility functions by introducing a discount factor, which reduces the present value of future payoffs, as a function of how far removed they are in time.

Cowen rejects this (offering instead a commitment to the “overtaking criterion” (128) as a way of comparing utility streams). Following Parfit (1986, 356-357), he draws an analogy between space and time, arguing that where a particular harm occurs is as irrelevant as when it occurs, and so temporal distance is as irrelevant as spatial distance when it comes to the moral assessment of the action that produces it. As Cowen puts it “we can think of the universe as a block of four-dimensional space-time. We would not discount human well-being for temporal distance per se any more than we would discount well-being for spatial location per se” (p. 69). One can see the influence of Peter Singer’s arguments in the analogy to space. Indeed, one way of thinking about the zero-discounting position is to see it as essentially a Singer-style ‘expanding the circle’ argument, which claims that the status we assign to our contemporaries is arbitrary, from the moral point of view, and so we must expand our moral horizon to include all future people as well.

While I had previously put it down as something of a curiosity that Cowen held this view, Stubborn Attachments shows that it is in fact quite central to his thinking across a range of different topics. Indeed, it is implicit in his discussion of economic growth, in which the key idea is that favoring the Crusonia plant will eventually pay off. It may not pay off for the current generation, or the next, but eventually people will be better off. This is why, on Cowen’s view, we must have stubborn commitments. We must be prepared to ignore the clamouring of people here and now, in order to pursue the policies that will be maximally beneficial in the long term.

The interesting suggestion that Cowen makes—the one that purports to resolve the fundamental tension in internet libertarianism—is that consequentialism, when combined with a commitment to temporal neutrality, allows one to derive a system of rights that function like strict deontic prohibitions in the present. Cowen points out that, given our epistemic limitations, especially given the complexity of causal influences on future events, our commitment to maximizing good consequences in the long term is going to have to translate into the adoption of standing policies. This is the standard rule-utilitarian move. However, as he points out, if one is concerned about all future
consequences, then the rights that institutionalize the Crusonia plant are going to function as exceptionless constraints in the present, because the sum of good consequences that they generate in the future is so great that they easily outweigh any countervailing considerations in the present. Thus the only grounds one could have for violating rights is the expectation that a failure to do so would somehow threaten human civilization.

In this way, Cowen attempts to pull a deontological rabbit out of a consequentialist hat. The apparent tension between the two forms of moral reasoning arises only because of our tendency to ignore the very distant future—something that, admittedly, most of us do. This is clearly the most interesting idea in Cowen’s book, although it is not particularly well-developed, and frankly, I am not sure that it can do the work that he expects it to do. At the same time, the argument is both original and provocative.

I will leave it to others to see if the details can be worked out, because I found myself objecting to features of the argument that arose long before this point. First, the idea that we can ‘grow our way out of class conflict’, which is central to Cowen’s view of the importance of Crusonia plants, is subject to much more serious objections than he acknowledges. One of the big bets made by Marx was that, as industrial production began to lift the curse of Adam, and people became increasingly wealthy, they would also become less concerned about the distribution of wealth. This is what made it possible to imagine a society entering ‘post-scarcity’ conditions, in which each would happily contribute according to his ability, and so forth. This bet is one that, famously, failed to pay off—average wealth has increased ten-fold, and yet individuals are not even a tiny bit less possessive than they were a century ago. There are various accounts for why this is, but the most important involves the suggestion that, as people become wealthier, an increased fraction of their consumption becomes instrumental to the attainment of status, which is essentially zero-sum.

One of the most surprising features of Cowen’s book is that, given his normative framework, he is not more worried about status competition. Thorstein Veblen and Fred Hirsch make no appearance in his discussion. Cowen deals with some concerns about status in a section dedicated to the claim that increased growth, beyond a certain point, fails to promote happiness, but he says nothing at all about the concern that growth does nothing to diminish distributive conflict (and
thus does nothing to resolve ‘aggregation problems’). This is especially peculiar given that, in his contributions as a public intellectual, Cowen often claims that the arguments being traded back and forth by other commentators, as well as the political positions taken, are merely disguised attempts to raise or lower the status of different groups. (Indeed, his well-known heuristic for reading op-ed articles is that most can be reduced to the claim that one group, with whom the author affiliates, should see its status rise, while some other group, with whom the author disassociates, should see its status fall.) Cowen has even argued that the literature on growth and happiness should be interpreted as an attempt by intellectuals to raise their relative status: “When happiness research indicates that money brings more happiness only up to a point, this is a popular result. That perspective lowers the status of this monied class by showing they really aren’t that happy” (Cowen, 2008).

And yet, if intellectual life is all just a disguised form of status competition, how hard is it to believe that economic life—once the basic necessities have been met—is also just an indirect or disguised form of status competition? And how hard is it to believe that hundreds of years in the future, consumption will have become entirely instrumental to relative status concerns? But if this is true, then the Crusonia plant of economic growth is not the open-ended source of gain that Cowen imagines it to be, nor can we have any reasonable expectation that it will allow us to avoid dealing with questions of distributive justice. Consider Pierre Bourdieu’s claim that aesthetic value is merely a disguised form of status seeking (Bourdieu 1984). I have no doubt that Cowen rejects this claim—he appears to have almost endearingly non-cynical views on aesthetic taste—but he should at least be worried that it is true, since his normative position is one that relies so heavily upon the positive-sum character of social life as a way of circumventing difficult philosophical questions.

I was also unpersuaded by Cowen’s commitment to temporal neutrality, along with his conviction that we must ‘expand the circle’ to include all future people. This is a very complex issue, which I cannot get into the details of here.6 There is not much of an argument in the book, other than the bald assertion that time is just like space, both irrelevant from the moral point of view. It is worth noting, however, that

---

6 For discussion, see Heath (2017).
Cowen’s formal derivation of temporal neutrality, in his paper “Consequentialism Implies a Zero Intergenerational Rate of Discounting”, relies upon the most objectionable feature of consequentialism to produce that result. In this paper, he introduces an axiom of “person neutrality”, which states that “a benefit for one person is equally good as the same benefit for another person” (Cowen 1992, 163). He then claims that “neutrality across generations becomes a natural extension of the traditional consequentialist belief in person neutrality within a single generation” (166). This idea—that one can simply take away happiness from one person and give it to someone else, and that these redistributions are all a matter of indifference from the moral point of view—is what critics like Rawls have in mind when they say that consequentialism fails to respect the distinctness of persons.

More generally, the commitment to temporal neutrality helps to explain a noticeable peculiarity in Cowen's public commentary, a trait that some have described as a lack of empathy. As a long-time reader of Cowen's work, I have also found it difficult on occasion to see how a person who is so concerned about promoting welfare could exhibit so little concern for actual human suffering. There is something rather Olympian, for instance, about Cowen’s insistence that, as a society, we spend far too many of our health care dollars caring for sick people, and not nearly enough investing in medical research. He often comes across like the proverbial socialist who loves the people, just not any specific person. Reading Stubborn Attachments has helped me to understand the underlying rationale for this stance. His zero-discounting principle is one that allows him to care deeply about humanity, without caring too much about actual people—i.e. the small subset of persons who happen to be alive at the moment. On Cowen’s view of things, caring about the suffering of real, live human beings shows up as a prejudice, since we are arbitrarily privileging the interests of those who occupy a block of space-time contiguous to our own.

And here, finally, we come to one of the most important differences between philosophy and economics. In philosophy, what one man considers a conclusion will more often be regarded, by another, as a reductio.
REFERENCES


Joseph Heath is Professor in the Department of Philosophy and the Munk School of Global Affairs and Public Policy at the University of Toronto. A fellow of the Royal Society of Canada and the Trudeau Foundation, Heath is the author of several books, both popular and academic. His most recent, *Morality, Competition and the Firm* (Oxford, 2014), is a collection of papers on business ethics and the normative foundations of market economies. He is also the author of *Enlightenment 2.0* (HarperCollins, 2014) and, with Andrew Potter, *The Rebel Sell* (HarperCollins, 2004).

Contact e-mail: <joseph.heath@utoronto.ca>
Website: <chass.utoronto.ca/~jheath>

ERIK SCHOKKAERT
*KU Leuven*

Towards the end of her stimulating and thought-provoking book, Kate Raworth summarizes her main message as follows:

The twenty-first-century task is clear: to create economies that promote human prosperity in a flourishing web of life, so that we can thrive in balance within the Doughnut’s safe and just space. It starts with recognizing that every economy – local to global – is embedded within society and within the living world. It also means recognizing that the household, the commons, the market and the state can all be effective means of provisioning for our many needs and wants, and they tend to work best when they work together. By deepening our understanding of human nature we can create institutions and incentives that reinforce our social reciprocity and other-regarding values, rather than undermine them. Once we accept the economy’s inherent complexity, we can shape its ever-evolving dynamics through smart stewardship. That opens up the possibility of turning today’s divisive and degenerative economies into ones that are distributive and regenerative by design. And it invites us to become agnostic about growth, creating economies that enable us to thrive, whether or not they are growing. (287)

This quote refers to the seven new ways of thinking about economic reality and about economic science that are developed in the book. The first way is the need to change the goal of the economy: rather than being fixated on GDP as a measure of social progress, humanity should focus on creating a social foundation of well-being that no one should fall below and an ecological ceiling of planetary pressure that the economy should not go beyond. Between the social foundation and the planetary ceiling lies the safe and just space for humanity, which she conceives to be a regenerative and redistributive economy. This vision of a better world is represented by the image of what looks like a doughnut.
Inspired by this vision, Raworth then puts forward six additional ways of thinking like a twenty-first-century economist (as she calls it): (ii) to go beyond the market and understand the important role played by the state, the households, the commons; (iii) to realize that economic agents are not only driven by narrow economic self-interest, but are social and interdependent human beings; (iv) to take into account that the economy should not be modeled as a system in mechanical equilibrium, but as a complex dynamic system with feedback loops and tipping points; (v) to see economic inequality not as an unavoidable feature of the economic system, but as a design failure that can be repaired; (vi) to accept that the ecological boundaries impose on us the need to build a regenerative circular economy; and (vii) to discover how the economy can overcome its addiction to growth.

I agree with Raworth's core message as summarized by the seven points above. I am firmly at her side when she describes what a good society consists in. I also agree of course that there is a deep gulf between this vision of how a good world should look like, and the present situation of the real world. She provides compelling arguments to show the absurdities that characterize our present economic system, i.e. the large and unacceptable inequalities, the exaggerated focus on material consumption (and GDP growth), and the lack of respect for the constraints imposed by our living planet.

From the point of view of economics as a science, the challenge raised by the doughnut economy is then to analyze the mechanisms that have brought humanity in the present situation, and the institutional changes that are necessary in order to bring us on a better development.
path. This general challenge indicates a number of more specific issues. For example, which measure of well-being should replace GDP? How should we manage the commons? How can we set up a democratic state that is not lobbied by big firms? How should institutions and incentive structures that reinforce our social reciprocity look? How should we organize an open source system without destroying incentives for innovation? How can we bring about more international coordination between sovereign countries in the face of climate change, and the obvious problems of free ridership in the actual negotiations? How should we design and implement a “global tax on extreme personal wealth” (200)? How can we guarantee to all world citizens a decent level of material consumption, while at the same time respecting the physical boundaries of the earth? I had hoped that the book would go deeply into these many difficult issues. Unfortunately, it does not.

As the quote in the introduction illustrates, Raworth does emphasize that ‘we’ can change the world, that ‘we’ should build a more democratic, and a more sustainable economy, that ‘we’ should be less motivated by GDP-growth. Yet, it is not always clear who ‘we’ is, and it is even less clear how ‘we’ should realize all this. To be fair, Raworth gives a number of inspiring examples of social experiments in different parts of the world; but these social experiments are limited in size and scope. This raises the question: why have these experiments remained so limited? How can ‘we’ move beyond the creation of a set of small beautiful islands in an ugly world? To answer this question, academics should analyze the basic mechanisms underlying our economic and social system. Raworth does not offer such an analysis. Rather, she suggests an explanation of sorts: the main cause of global injustices and environmental degradation is that we have bad economics. Mainstream economics is the culprit. One should note that the title of the book does not refer to the need for a doughnut economy, but to the need for doughnut economics.

This explanation is not very credible. I do not think that anybody really believes that the influence of economic ideas can be so big as to explain everything that goes wrong in the world. It is true that simplistic ideas may be very influential in social debates, and that flawed economic reasoning is used and misused by people in power, such as politicians, business persons, and climate deniers, to defend the status quo. It is therefore true that it is necessary to fight this flawed economics. However, my contention is that good mainstream economics
is one of the best instruments to correct these wrong arguments. The
target of Raworth’s criticism is not real mainstream economics, but a
caricature of it. To some extent her position is understandable as the
dominant view in many handbooks of economics unfortunately is close
to the caricature (I return to this point below).

One of Raworth’s tricks to denounce the mainstream is to define all
interesting economists as non-mainstream (even if they are affiliated to
traditional top economics departments and have been awarded the
Nobel Memorial Prize in Economics like Kahneman, Sen, Deaton, Thaler,
Stiglitz, and Ostrom). Some of the contributions she describes as
influential (the flow model in macroeconomics, the stages of Rostow)
were explained to me as a student in the 1970’s, and already then
presented as relics from the past—and her claim that the simple model
of rational economic man is not questioned by mainstream economists
nowadays is utterly surprising for everybody who really follows the
recent developments in economics. Raworth is very keen to applaud the
fact that critical students from the Kick Over movement plastered
accusatory posters in the Boston Hotel where the “highbrow” (her
formulation, 3) conference of the American Economic Association’s
annual 2015 meeting took place. According to Raworth, these posters
denounced the “gulf between the preoccupations of mainstream
economic theory and growing real-world crises such as global inequality
and climate change” (1). But did she check what was on the program of
that conference? The proceedings (published in the ‘highbrow’ American
Economic Review of May 2015) show that the keynote by Raj Chetty was
on “Behavioral Economics and Public Policy”, and that there were
sessions devoted to Thomas Piketty’s *Capital in the Twenty-First
Century*; impacts of the great recession on low-income households; high
stakes energy and environmental problems in developing countries;
empirical climate economics; culture, trust, and productivity; moral
values and economic behavior; behavioral economics in the classroom;
social behavior in the field; measuring the multinational economy;
education, race, and underrepresented minorities; the economics of
gender. Apparently, many economists do attempt to bridge the gulf
between economic theory and real-world issues.

It may look overly defensive and even a bit childish for a mainstream
economist to complain about Raworth’s biased presentation of what
mainstream economists are doing. Yet, I will argue that there is indeed a
deep problem here.
First, one can reasonably defend the position that, while it is silly to deny that economists are concerned with real world problems, there are still important flaws in economic methodology. Importantly, most economists give scant attention to natural resource constraints. There is certainly room for a serious debate on the strengths and limitations of methodological individualism, related to a discussion about the pros and cons of complexity dynamics (without a well specified behavioral model). However, nothing of this is found in the book. From the point of view of changing economics itself, constructing and attacking a straw man is not at all productive.

Second, and perhaps more relevant to Raworth’s purpose, there is the important question of why primitive economic ideas about the optimality of markets and the need of GDP growth (or, even worse, the use of GDP as an indicator of well-being) have become so influential in the real world. Why are activists like Raworth so often confronted with single-minded economists objecting with overly simplistic arguments to the points raised in the citation with which I started this review? In fact, why are the basic handbooks of economics still dominated by these simple ideas? Of course, the debate within the discipline to some extent reflects ideological differences, and economists in general are likely to have a conservative bias. I understand and share Raworth’s frustrations and her strong feelings about arrogant and myopic economists in power. However, I think that there is a more fundamental cause underlying this phenomenon.

In my view, already from its very beginning, economics has always been characterized by a confusing mixture of normative and positive considerations. This was certainly true for the classical economists, for the Marxians, for the marginal revolution, for Keynesianism, and for the new classical macroeconomics. It is a fascinating intellectual question how an apparently unregulated market economy does not collapse. It is therefore interesting to look at the features of what is called a general equilibrium. Yet, it is a further step to claim that under a set of (unrealistic) assumptions that general equilibrium is Pareto-optimal. And, it is an even further step to forget about the unrealistic assumptions that give rise to the a priori conviction that the unregulated market outcomes are socially desirable. This approach begs deep normative questions on the status of individual preferences as a criterion for the evaluation of economic outcomes. Moreover, it leads to the view that externalities are the exception rather than the rule, and
that intervening in the market leads to harmful ‘distortions’. It is this set of convictions that is still reflected in many (most) standard principles courses. And it is this set of convictions that indeed should be questioned—as Raworth does.

Yet, the research frontier in economics has obviously moved beyond these simple statements. The world of mainstream economics is now a world of behavioral agents in imperfect market settings who are influenced by non-monetary considerations. In fact, one of the good developments in recent mainstream economics is that it has become more empirically oriented. Of course, Raworth is right that the conservative bias is still strong; but, the economic toolbox has become much richer than it used to be in the past, and it can now be very helpful to better understand the mechanisms that are at work in the real world.

My main problem with Raworth’s book is that it suffers from the same deep problem that has plagued economics from the beginning: an unfortunate and confusing mixture of positive and normative considerations. Most of her policy prescriptions are based on beliefs, hopes, and convictions, rather than on serious empirical analysis. Rather than giving new insights, she proposes a set of beliefs to counter another set of beliefs. Raworth’s book is an important contribution to the ideological struggles about what is a good economy. It is a convincing competitor for handbook economics; but, despite its title it is not very helpful for those who want to build an attractive empirically-oriented doughnut economics.

I think this is highly problematic—both for handbook and for doughnut economics. Let me return to the issues mentioned at the start of this review. It is true that individuals may be altruistically motivated, that they are not exclusively motivated by financial self-interest, and that reciprocity is an important motivation—but, it would be very dangerous to assume that this is always the case and to completely neglect the presence of self-interested free riders. It is true that the power of the commons has been underrated by many economists—but, serious questions remain about the way to handle these commons. It is true that the traditional anti-state and pro-market bias of many economists is a very unfruitful attitude in the light of the great challenges before us—but, the state can also be an instrument of the powerful and the rich to exploit the poor, and the design of democratic institutions is not straightforward, certainly not at the global level. It is
true that economic inequality in the present world is unacceptable, and that it is essential to shift attention from purely redistributive measures to the mechanisms determining the primary distribution of income—but, much more work is needed to better understand the pitfalls in the move to such a more democratic economy. It is true that we should steer innovation and research so that they can contribute to a more just, and a more durable economy—but, steering innovation and research is far from trivial. And all these challenges become even more acute when we fully take up the resource constraints facing humankind.

If we want a better world, we certainly have to go beyond the market and growth ideology of many economists now. But it is equally important to go beyond simple activist beliefs in a new world. Activism is important, but it may also be dangerous. I fully side with Raworth’s claims about what would be a good economy. I think however that the transition towards such a good economy will not be easy and that detailed design issues are crucially important. Mainstream economics certainly will not give the one and only definitive answer, but it offers a coherent framework to tackle some of the main problems with the transition. Rather than construct a ‘mainstream’ strawman and kill it, it would be far more reasonable to try to establish a coalition with real ‘mainstream’ economists who agree with her basic views: that the model of the market has to be embedded in a broader view of society, that the economy should take into account resource constraints, that the distribution of economic resources reflects to some extent inequalities in power, that human beings are not only calculators driven by economic self-interest, but have broader values in life and imperfect decision-making capacities. All these perspectives are present in mainstream economics.

Raworth’s extremely well-written book has been influential in proposing the idea of the doughnut economy. That is fine: the more people are convinced that the economy has to be put on a different development path, the better. It is also effective in showing that the simplistic conservative interpretation, and abuse of some economic insights, is dangerous. That is another good thing. Yet, it aims at the wrong target by discarding all insights from modern mainstream economics—in fact, it simply does not mention any of these insights. At that point, it becomes more of a sermon than an analysis. This is a missed opportunity.
Erik Schokkaert is full professor of welfare economics at the Department of Economics of the KULeuven. He has been visiting professor at the London School of Economics, the Université catholique de Louvain and the Universidad ICESI in Cali. His research focuses on (i) the modelling of different notions of distributive justice; (ii) the concept of individual well-being and quality of life; (iii) the application of these concepts to the analysis of specific policy problems in the fields of health, social security, and taxation. He has published in (among others) *American Economic Journal: Microeconomics, Demography, Rand Journal of Economics, Journal of Health Economics, Health Economics, Journal of Public Economics, Social Choice and Welfare, Economics and Philosophy, European Economic Review, Health Policy,* and *Economica.* Together with Wulf Gaertner he has written a book on *Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice* (Cambridge University Press, 2012).

Contact e-mail: <erik.schokkaert@kuleuven.be>
PHD THESIS SUMMARY:
Understanding with Models

PHILIPPE VERREAULT-JULIEN

Doctorate in philosophy, January 2019
EIPE, Erasmus University Rotterdam

Theoretical modelling raises many epistemological challenges. Models appear to explain phenomena, yet it is not obvious that they satisfy requirements that typical theories of scientific explanation deem essential, for instance faithfully representing causal factors of interest (Reiss 2012). My doctoral thesis aims at answering the following central and general question: What, if anything, can we learn from theoretical models? The overarching project is to show that models, especially in economics but also in other disciplines, may provide epistemic benefits in the form of understanding even when they do not actually explain or when they do not provide causal knowledge.

Chapter 2 analyses what I call the inferentialist-behavioural account of understanding (InfBUn) (Ylikoski 2009; Ylikoski and Aydinonat 2014; Kuorikoski and Ylikoski 2015). According to InfBUn, what constitutes understanding is having knowledge of relations of counterfactual dependence and having the ability to draw what-if inferences. Drawing on literature in epistemology and philosophy of science, in this chapter I argue that InfBUn is inadequate. First, I show that the behavioural concept of understanding cannot properly distinguish illusory from genuine cases of understanding. This is because it places excessive emphasis on actual inferential performance as a criterion for understanding rather than the ability proper. Second, I contend that it is not necessary to have a behavioural concept of understanding in order to retain compatibility with the view of models as extended cognition. I then propose that InfBUn should be better viewed as an evaluative account of understanding and not as a substantive one.

The third chapter examines what I call the factive inferentialist account of representation (FInfR) (Kuorikoski and Ylikoski 2015; Kuorikoski and Lehtinen 2009). FInfR has two key features. First, it holds that models represent in virtue of the correct what-if inferences they afford. Second, FInfR claims that, contrary to the traditional
inferentialist account of representation (Suárez 2004), it is compatible with the factivity of explanation, i.e. the requirement that explanations have a true explanans. I argue that FlnfR does not, in fact, allow to demarcate merely phenomenological models from explanatory ones. This is because the notion of correct what-if inferences is either too ambiguous or too liberal. I then expound a programmatic dilemma FlnfR faces. The first option is to double down on deflationism, viz. dismissing any substantive criteria about successful representation. The second one is instead to substantiate under what conditions what-if inferences are explanatory. I argue that the first horn is undesirable because it implies abandoning FlnfR’s main motivation, namely its commitment to the factivity of explanation and to realism.

Chapter 4 (part of it is published in Verreault-Julien 2019b) first argues against two sets of views one can commonly find in the literature about the relationship between understanding and explanation:

1. Causal knowledge is necessary for understanding.
2. Only explanations can provide that knowledge.

Following a terminology I borrow from Pritchard (2014), I call the conjunction of these views, or tenets, the narrow knowledge account of understanding (narrow KAU). This chapter has three aims. The first is to show that narrow KAU faces descriptive and normative issues because it cannot account for scientific practices that do not actually explain (e.g., theoretical modelling) and for those that do not explain causally (e.g., mathematical explanations). The second aim is to debunk the tenets of narrow KAU. The third aim is to propose an alternative to narrow KAU, one that I coin broad KAU. Broad KAU, contrary to its narrow sibling, makes the following two propositions:

1. Causal knowledge is not necessary for understanding.
2. Having an (actual) explanation is not necessary for understanding.

I propose an account of broad KAU that builds on Reutlinger’s (2016) counterfactual theory of explanation. I show that, by amending the veridicality condition of his theory in favour of what I call the possibility condition, we are left with an account of understanding. While the veridicality condition requires the actual truth of the explanans and the explanandum, the possibility condition only demands their possible truth. This allows to separate explanation from understanding.
Explanation requires to have true information about actuality whereas understanding only requires true information about possibility. This account, I contend, can accommodate both mathematical and how-possibly explanations.

In chapter 5 (separately published in Verreault-Julien 2019a), I propose and develop a new account of how-possibly explanations (HPEs). HPEs are usually contrasted to how-actually explanations (HAEs). Instead of taking the degree of empirical support as the demarcation criterion between HPEs and HAEs (e.g. Bokulich 2014), I suggest that the fundamental feature of HPEs is that they provide knowledge of possibility. While HAEs are characterized as propositions of the form ‘p because q’, HPEs include a possibility operator and thus have the form ‘(p because q)’. Crucially, HPEs have a truth value, they can be true or false. Evaluating their truth value requires at least a minimal amount of relevant background knowledge. Finally, I argue that models should not be viewed stricto sensu as HPEs. Rather, models provide reasons to believe propositions of the form ‘(p because q)’.

Chapter 6 (also published in Verreault-Julien 2017) uses many insights from the previous chapters and applies them to a specific case in economics, namely the Arrow and Debreu (1954) model of general equilibrium. The model is considered to be a landmark achievement of economic theory, yet it is also often criticized (Blaug 2002). The main claims of this chapter are that the general equilibrium model provides a mathematical HPE and that this HPE affords understanding not only of the model, but also of the world. I show that proving the existence of a general equilibrium was motivated by the fact that it could inform economists, albeit minimally, about its actual existence. This is because mathematical claims can serve an evidential role with respect to causal claims. For if the equilibrium would be mathematically impossible, then this would be evidence that it is causally impossible.

REFERENCES

---

1 This reads ‘It is possible that (p because q)’.


Philippe Verreault-Julien obtained his PhD in January 2019 from Erasmus University Rotterdam under the supervision of Julian Reiss (Durham University) and Jack Vromen (Erasmus University Rotterdam). His research mainly focuses on the epistemology of theoretical modelling. Starting fall 2019, he will be a postdoctoral researcher at the Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science. He acknowledges the financial support of the Fonds de recherche du Québec – Société et culture (FRQSC) for his PhD research.

Contact e-mail: <verreaultjulien@esphil.eur.nl>
Website: <https://pvjulien.net>
PHD THESIS SUMMARY:

JAMES GRAYOT
Doctorate in philosophy, February 2019
EIPE, Erasmus University Rotterdam

In this thesis, I offer a philosophical perspective on the different conceptions of agency and choice as they are understood and employed in economics and behavioral decision research—this perspective is two-fold: on the one hand, philosophical analysis can clarify ambiguities in definitions and concepts that arise within interdisciplinary research. This is of particular importance given how philosophical concepts such as mind, cognition, and intentionality feature in economic studies of rational choice. Hence, one project of this thesis is to subject contemporary research on questions about agency and choice to such philosophical scrutiny. On the other hand, the questions and topics discussed in this thesis can be understood as an exercise in philosophy of science: they deal explicitly with questions and topics that pertain to the theoretical and empirical practices of scientists. This includes traditional microeconomic disciplines, such as decision theory and game theory, as well as interdisciplinary research in behavioral economics, neuroeconomics, and experimental psychology.

Chapter 2 illustrates the contentious relationship between agency and choice by focusing on a cluster of debates in the philosophy of economics about the scientific status of preferences. These debates revolve around a central question, namely, whether preferences should be construed as mental or behavioral phenomena. The literature indicates that these debates are critical to the disciplinary identity of economics (Hausman 1998; Davis 2006; Bruni and Sugden 2007; Backhouse and Medema 2009; Hands 2009, 2014; Ross 2011, 2014), and further, that disputes about the role of psychological explanations in economics are predicated on how one conceives economics as a science (Dowding 2002; Camerer 2008; Hausman 2008, 2012; Guala 2012, 2017; Gul & Pesendorfer 2008; Hands 2013; Clarke 2016; Okasha 2016;
Dietrich and List 2016). Chapter 2 thus evaluates the relevance of the mentalism-behaviorism dichotomy in economics in light of these debates. In particular, I argue that there are two problems with current conceptions of the dichotomy as it pertains to how economists and decision researchers interpret and gather evidence. First, it is unclear what the dichotomy pertains to or is about exactly—which is to say, economists and decision researchers may have different motivations for endorsing mentalism and/or for opposing behaviorism. Second, and more importantly, it is unclear how the mentalism-behaviorism dichotomy is supposed to improve or advance empirical research in economics—in particular, supporters of mentalism have the difficult task of clarifying what mentalism entails or consists in. In response to the first problem, I consider two common motivations for endorsing mentalism: one motivation appeals to the choice-theoretic foundations of economics; the other appeals to scientific practice in economics. In response to the second problem, I argue that the mentalism-behaviorism dichotomy is not likely to advance or improve scientific practice in contemporary economic settings because neither mentalism nor behaviorism are equipped to analyze and resolve explanatory problems that are specific to the interpretation of psychological and neuroscientific (i.e. non-choice) data. I conclude by discussing the limitations of functionalism, the mainstay of the mentalism defense book, and suggest alternative schemas to the mentalism-behaviorism dichotomy, some of which are employed in neighboring areas of the cognitive and behavioral sciences.

Chapter 3 considers whether, i.e. under what conditions, human persons behave like economic agents. It is now recognized that humans are boundedly rational, which means that persons typically do not think and behave like *homo economicus* agents. Among the methodologies for modeling boundedly rational individuals, multiple-self models have gained considerable popularity as tools for representing the dynamics of intrapersonal choice under various conditions and constraints. Multiple-self models typically work by isolating features endogenous to individuals that motivate them to act in different ways. Generally, these features are taken to correspond to autonomous structures *within* the individual and, as such, are modeled as if they were independent agents (that is, independent agents who can reason together). In contrast to debates discussed in chapter 2, which take an individualistic perspective toward the analysis economic concepts and decision phenomena, this
Chapter demonstrates how external forces such as social institutions and informational structures both support and constrain individual behaviors. I argue that individualism is problematic as a basis for investigating social interaction. In so doing I examine the Don Ross’ (2005, 2006) account of multiple-selves as a way of reconciling individuals’ bounded rationality with their bounded individuality. Ross argues that individual persons are complex aggregations of selves, which arise in response to external pressures to regulate individual behaviors—this enables the tracking of public norms and conventions. I thus investigate the different roles that ‘selves’ play in Ross’ broader philosophy of economics and I identify separate projects that arise therein. I distinguish three different roles for selves, which are evolutionary, narrative, and economic, and I argue that these roles contribute to two distinct, but overlapping, projects. I show that, while it is not problematic to conceive of selves according to their different roles, we should not presume that the functions or properties of selves in one role can serve the same purposes for different projects.

Chapter 4 returns to the domain of individual decision-making—it asks: How do interdisciplinary approaches to decision research integrate psychological insights with economic methods? And, what are the conceptual and ontological challenges of such integration? The idea of the ‘divided self’ has been the source of folk-wisdom for centuries. However, new research into the cognitive and behavioral foundations of decision-making suggests that this idea is more than just a metaphor. Our minds—and brains—appear to be divided in interesting if unexpected ways. Having elaborated the importance of external forces for understanding humans’ quasi-economic agency, Chapter 4 critically examines how multiple-self models of intrapersonal and intertemporal choice have been integrated with dual-process and dual-system theories from cognitive psychology. I adopt the term ‘multiple-agent model’ to denote a special kind of interdisciplinary model which conceives of multiple agents with contrasting psychological abilities. For example, Bénabou and Tirole (2002), Bernheim and Rangel (2004), Benhabib and Bisin (2005), Loewenstein and O'Donoghue (2005), and Fudenberg and Levine (2006) each have sought to characterize the contradictory tendencies of temporally distinct selves by investigating how controlled and automatic processes influence choice behaviors over time. In some instances, the intrapersonal dynamic between sequential selves is taken to establish the limitations on the decision-maker's ability to exhibit
self-control (Bénabou and Tirole 2002; Fudenberg and Levine 2006). In other instances, the conflict between an individual’s desire to consume now or later is interpreted as a trade-off between distinct systems, whose aims are regulated by the activation of different cognitive processes (Benhabib and Bisin 2005; Loewenstein and O'Donoghue 2005). Likewise, some neuroeconomic approaches to decision-making have modeled brain processes based on what economists interpret to be optimizing procedures. Research conducted by McClure et al. (2004) and McClure et al. (2007), and further results obtained by Brocas and Carrillo (2008a, 2008b, 2014), suggest that individual decisions are the outcome of strategic interactions between domain-specific systems. This technique presumes that the brain has limited energy resources and that it must allocate those resources efficiently in order to satisfy rewards. In this way, the brain is modeled as an optimizer with budget constraints. These models seem to be growing in popularity given their purported ability to predict and explain reasoning errors and decision anomalies due to internal conflict or lack of self-control. In chapter 4, I analyze how multiple-agent models employ the concepts ‘selves’ and ‘systems’ for the purposes of representing intrapersonal and intraneural conflict. Herein I defend three claims. The first and second claims argue that multiple-agent models are conceptually as well as ontologically ambiguous. The third claim argues that such ambiguities can lead to problems in scientific understanding. The examination of multiple-agent models is not only important for understanding how economists and psychologists jointly interpret and model self-control problems, but it further presents an important opportunity to study the effects of cross-disciplinary pollination of concepts and theories.

Chapter 5 builds on the analytical framework developed in Chapter 4 and investigates the role(s) that dual process theory has played as a psychological framework in behavioral economics and neuroeconomics. Cognitive scientists and philosophical psychologists alike have criticized the theoretical foundations of the standard view of dual process theory and have argued against the validity and relevance of evidence used to support it (Gigerenzer and Reiger 1996; Osman 2004; Keren and Schul 2009; Kruglanski and Gigerenzer 2011). Moreover, recent modifications of dual process theory in light of these criticisms have generated additional concerns regarding its applicability and irrefutability (Keren 2013; Mugg 2016; Pennycook 2017; Bonnefon 2018). I argue that this should raise concerns for behavioral economists and neuroeconomists.
who see dual process theory as providing psychologically realistic foundations for their models. In particular, it raises the possibility that dualistic models are not as descriptively accurate or reliable as economists presume them to be. In fact, the case can be made that the popularity of dual process theory in economic decision research has less to do with the empirical success of dualistic models, and more to do with the convenience that the dualistic narrative provides economists looking to sort out decision anomalies (cf. Grüne-Yanoff 2017). I argue that the growing number of criticisms of dual process theory leaves economists with something of a dilemma: either they stick to their purported ambitions to give a realistic description of human decision-making, or they stick to dual process theory and revise their scientific ambitions.

In Chapter 6, I offer concluding remarks and consider where one goes from here. Chapters 2–5 project two main approaches to reconciling the tension between agency and choice. One approach views individual persons as the primary objects of study for economics, and as such, indicates that psychology and neuroscience can help improve how economists study rational choice. The second approach views individual persons not as the primary object of study (economic agents are the primary object of study, and they are ontologically distinct from persons). As such, choice should be construed as the outcome of external (market) pressures, which include important socio-cognitive supports. Hence, for each of these approaches, there are new pursuits and new philosophical questions to be considered.

REFERENCES


James Grayot completed a PhD and research master in philosophy at the Erasmus Institute for Philosophy & Economics (EIPE) at the Erasmus School of Philosophy. He does research on philosophy of science, working on issues at the intersection of economics, psychology, and cognitive (neuro)science. He also works on philosophy of mind, philosophy of action, and decision theory. He is presently a visiting fellow at the Tilburg Center for Logic, Ethics, and Philosophy of Science.
(TiLPS), and a lecturer at the Vrije University Amsterdam, and the University of Groningen.

Contact e-mail: <james.grayot@gmail.com>
This thesis studies the way in which economists have sought to establish the scientific authority of their discipline during the period before and after World War II in the United States. The research shows how the quest for scientific authority by economists gave rise to new concepts and notions, instruments of control, and calculation methods. Such developments contributed material and symbolic advantages to the discipline in the academic world and the broader academic sphere.

The main contribution of the thesis is its consideration of the heterogeneity of scientific approaches used by economists in different institutions. To capture this heterogeneity, and to highlight the practical and political dimension of epistemological issues regarding the definition of science, the research was guided by the notion of *boundary-work* (Gieryn 1983, 1995, 1999). Similarly, the application of Pierre Bourdieu's sociology (Bourdieu 1988, 1998, 2004; Bourdieu and Wacquant 1992) made the clarification of interactions between the academic world and other social spheres possible. The thesis is based on a significant amount of secondary literature and the consultation of primary sources (The Economists' Papers Archive at Duke University, and the AEA online archives).

The analysis focuses on three of the institutions at the pinnacle of the discipline in the American academic world: the Cowles Commission, the Economics department of the Massachusetts Institute of Technology, and the Economics department of the University of Chicago. The thesis is divided into three parts. Part one (chapters 1 to 3) highlights the *specific inflections* of each institution based on its institutional history and the profile of its researchers. Part two (chapters 4 to 6) studies the strategies put in place in each site to establish the scientific authority of the discipline. The third part (chapter 7) sheds light on the elements that, after World War II, combined to create the
conditions for development of the disciplinary consensus on which economists frequently base their claims of scientific quality.

At the Cowles Commission, what is considered scientific was connected to a form of abstract knowledge whose objective was universality. The emergence and crystallisation of this concept of scientific knowledge are examined in three episodes: the preambles of the constitution of the Econometric Society (ES) (1930–1932), the controversy of the ‘Mesure sans théorie’ (CMST) (1947–1949), and the publication of *Three Essays on the State of Economic Science* by Tjalling Koopmans (1957). By closely observing these three different episodes, the thesis first describes how the network of researchers associated with the Cowles Commission was able to pass from a fringe position to a central and dominant one within the discipline. Secondly, it explores the conditions that progressively led to economics articulating a response to the criticism that was calling into question the focus on formal and mathematical problems per se.

At the Economics department of the Massachusetts Institute of Technology, scientific knowledge was assimilated with a form of technical knowledge. The thesis shows how, within the framework of an engineering school closely connected to the political and military powers, a *translation* between ‘fundamental’ research and ‘concrete tools’ was undertaken. To explain the conditions that made the development of this kind of knowledge possible, the thesis focuses on the pressures which demanded effective use of resources, exacerbated especially during World War II, and on the changes in the organisation and production of knowledge of the academic system of the United States during the postwar period. In concrete terms, the thesis shows how, by placing the laboratories at the heart of research and pedagogy, the administrators of MIT effectively interconnected the academic world and the practical sphere.

At the Economics department of the University of Chicago, science was associated with a form of empirical knowledge founded on statistical data and tested through predictive accuracy. The research shows how, through this association, economics established its qualifications as an applied discipline heavily involved in political decision-making. In particular, the classification of economics into ‘positive means’ and ‘normative ends’, as proposed by Milton Friedman, enabled the discipline to put forward a methodological response to the problem of the role that economists could play in the process of
political decision-making without losing their scientific authority. In order to study this process, the thesis analyses the materialisation in the department’s practices of Friedman’s methodological precepts, his version of the ‘theory of prices’ and, more broadly, his positivist convictions.

The standardisation of the doctorate degree is presented as the central element of the development of the disciplinary consensus. To assess this, the thesis is based on the ‘Bowen Report’ (Bowen 1953). The thesis shows how, based on the standardisation of the doctorate degree, economics has defined a professional and intellectual locus of authority while simultaneously developing a relatively unified body of knowledge designed to preserve a consensus of thought.¹

Considered as a whole, the processes studied in the thesis provide two distinguishable but interconnected scales of analysis. The first is local, meaning that it concerns the dynamics suited to each of the three institutional sites studied. The second is general and studies the consequences for economics as a discipline of processes studied on the scale of each institution. One of the main results of the thesis is therefore the construction of a disciplinary matrix that connects the two scales of local and overall analysis. Firstly, it sheds light on the fact that the increasingly unanimous acceptance of the expression ‘mainstream’ as a practical label does not however imply the existence of widespread agreement on the meaning of this expression.

Secondly, the connection of the results from the local scale within a disciplinary matrix has made it possible to capture the importance of the existence of a mechanism of accumulation of knowledge for the discipline’s internal cohesion. In more concrete terms, the thesis identifies three mechanisms of accumulation of different knowledges in economics. To Cowles, the axiomatic method proposed by Koopmans in the Three Essays built on the guidelines of a mechanism whose mainspring was the construction of models that, progressively, were designed to take into account various aspects of markets and economic behaviour. Rather than rendering the hypotheses flexible so that they might take stock of aspects of reality, the accumulation of knowledges at Chicago operated via the broadening of the range of their applications. This method of accumulation of knowledges enabled

¹ An article presenting the main results of this chapter has been already published in French: see Orozco Espinel (2017).
economics to broaden the field of application of the discipline and led economists onto the terrain of other fields—notably sociology, law, and political science. The process of accumulation of knowledge at MIT operated very differently. The way in which the economists of MIT represent economic behaviour through simple mathematical models—aimed initially at understanding only a few aspects of a particular situation, to then apply those to a broad range of problems—served as a vector for the accumulation of knowledges. Among the problems to which the mechanism can be applied, we find trade, consumer behaviour, and public spending.

Finally, the disciplinary matrix was central to explaining why the production of economists with varied interests, abilities, and skill sets was initially possible and later proved very useful for enhancing the social authority of economics as a whole.

The disciplinary matrix also has heuristic potential. While the study focuses on three sites, the proposed analysis—focusing more on the discipline of economics—allows contributing to the wider problem of taking heterogeneity into account. Concretely, the analytical device established in the research opens up the possibility of articulating processes present in other institutions, notably those that are not highly ranked within the discipline, and thus also enabling the compression of the history of the discipline to be expanded, through the recognition of its richly diverse and iridescent character.

REFERENCES


**Camila Orozco Espinel** is an interdisciplinary scholar. She obtained her PhD in Sciences of Society at the EHESS (École des Hautes Études en Sciences Sociales, School of Advanced Studies in the Social Sciences). Her research focuses on 20th Century history of economics, primarily the post-Second World War quest of the discipline for scientific authority. She also pursues research on women economists in Latin America, the contribution of women to econometrics, the collaborations between couples of economists and history of the International Association for Feminist Economics. She teaches economics at the University of Lille, France.

Contact e-mail: <camilaorozco@gmail.com>