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, Erasmus School of Philosophy, Erasmus University Rotterdam. EJPE publishes research on the methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of inter-disciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>.

EDITORS
Savriël Dillingh
Gideon Frey
Elisabetta Gobbo
Ruth Korte
Benjamin Mullins
Erica Celine Yu

ADVISORY BOARD

EDITORIAL ASSISTANTS
James Grayot, David Holroyd
ARTICLES

The Partially Impartial Spectator
SVEINUNG SUNDFØR SIVERTSEN [pp. 1–24]

ARTICLE SYMPOSIUM on Anodyne Privatization

Anodyne Privatization
JOSEPH HEATH [pp. 25–65]

Privatization, Structural Dependence, and the Problem of Legitimacy
CHIARA CORDELLI [pp. 66–84]

Repoliticizing Privatization
SAVRIËL DILLINGH [pp. 85–106]

Elements for a Normative Theory of Privatization
RUTGER CLAASSEN [pp. 107–135]

Public Provision in Democratic Societies: Reasons to Reject Privatization
MARTIN O'NEILL [pp. 136–166]

INTERVIEWS

Ethics from the Outside Looking In:
An Interview with ROGER CRISP [pp. 167–200]

What Public Policy Can Be:
An Interview with MATTHEW ADLER [pp. 201–250]
BOOK REVIEWS

Olúfẹmi O Táiwò’s *Reconsidering Reparations*

JOSEP RECASENS

[pp. 251–254]

Josiah Ober's *The Greeks and the Rational: The Discovery of Practical Reason*

BOB R.W. VAN VELTHOVEN

[pp. 255–260]

Thomas Piketty’s *A Brief History of Equality*

ROSS EMMETT

[pp. 261–265]

André et al.’s *From Evolutionary Biology to Economics and Back: Parallels and Crossings between Economics and Evolution*

AHMED AL-JUHANY

[pp. 266–271]

Mark Fabian’s *A Theory of Subjective Wellbeing*

WILLEM VAN DER DEIJL

[pp. 272–279]

Thomas Kelly’s *Bias: A Philosophical Study*

LENNART B. ACKERMANS

[pp. 280–286]

PHD THESIS SUMMARIES

The Singular Plurality of Social Goods: Social Ontology and Collective Dilemmas

MARCO EMILIO

[pp. 287–291]

Defining Exploitation

ULYSSE LOJKINE

[pp. 292–295]

Sharing in Common: A Republican Defence of Group Ownership

YARA AL SALMAN

[pp. 296–299]
The Partially Impartial Spectator

SVEINUNG SUNDFØR SIVERTSEN
University of Bergen

Abstract: According to Adam Smith, we appeal to the imagined reactions of an ‘impartial spectator’ when justifying moral judgements of others and aspire to be impartial spectators when making judgements of ourselves. However, psychological research has shown that trying to be impartial will often have the paradoxical effect of reinforcing other-directed prejudice and self-serving bias. I argue that we can get around this problem by aspiring to be ‘partially impartial spectators’ instead.

Keywords: Adam Smith; impartial spectator; partial impartiality; ethical ideal; naïve realism; fundamental attribution error; bias; motivated reasoning; accuracy goal; directional goal

JEL Classification: B120, B310, B400

We endeavour to examine our own conduct as we imagine any fair and impartial spectator would examine it.

— Adam Smith, The Theory of Moral Sentiments (III.1.2, 129)¹

“You listen to me. This is one time you're not going to get away with twisting everything I say. This just happens to be one damn time I know I'm not in the wrong. You know what you are when you're like this?” “Oh God, if only you’d stayed home tonight.” “You know what you are when you're like this? You're sick. I really mean that.” “And do you know what you are?” Her eyes raked him up and down. “You're disgusting.”

— Richard Yates, Revolutionary Road (2011, 35)

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 (in upper case Roman numerals), section (where applicable, in lower case Roman numerals), chapter, and paragraph (in Arabic numerals).

AUTHOR’S NOTE: Thanks to Carola Freiin von Villiez, two anonymous reviewers, and the Editors at EJPE for their contributions and corrections.
If the two of us disagree, we may call upon a third person to help settle our conflict. Frequently, however, no third person is at hand, and we are left to our own devices. Fortunately, as Adam Smith notes, those devices include the ability to imagine what a third person, a spectator, would think, if present.

Unfortunately, as Richard Yates’ *Revolutionary Road* reminds us, it is rarely easy for two people in conflict to agree what the reaction of this imaginary spectator would be.

This is true even if we agree that the imaginary spectator should be both “well-informed and impartial” (TMS, III.2.32, 150). A large body of evidence from research on psychological bias indicates that the very effort to “see our conduct from without, as any impartial spectator would see it” (TMS, III.1.2, 129) will backfire. Rather than getting a more impartial impression of ourselves, we each end up convinced that “this just happens to be one damn time I know I’m not in the wrong”. Fast on the heels of this conviction comes the seemingly logical but disastrous conclusion that, if the other fails to agree, there must be something wrong with them: “You know what you are when you’re like this? You’re sick”.

It is not immediately apparent how we are to avoid this outcome. We might try harder to be well-informed and impartial, but, as we shall see, this also easily backfires, leaving us worse off than we were before making the extra effort. Instead, I will argue, we can make at least some headway on the problem by trying smarter. Specifically, we can modify the ideal of the “impartial and well-informed spectator” (TMS, III.1.2, 129) to account for some of the most pernicious and commonplace biases that distort our moral judgements, and then aim for this modified ideal instead. Before I go into my argument for why and how we should aim to be ‘partially impartial spectators’, let me give you a brief history of the idea and ideal of the impartial spectator as Smith conceived of it.

**I. THE IMPARTIAL SPECTATOR AS A MORAL ANCHOR**

One of the thickest threads running through the yarn of Adam Smith’s *The Theory of Moral Sentiments* is that of the impartial spectator. That epithet is sometimes applied to actual people who happen to be sufficiently “indifferent” to judge us impartially (TMS, I.ii.4.1, 47). Its most important referent, however, is the *idea* of such a spectator, playing the role of an ethical ideal in our moral judgements (I.ii.2.1 97).
In a sense, the idea of the impartial spectator arises out of frustration. Smith saw humans as fundamentally sociable, driven by “an original desire to please, and an original aversion to offend” other people (TMS, III.2.6, 135). In order to please and not offend, we “accustom ourselves” to consider our behavior from the perspective of other people (TMS, III.2.6.32, 152, n22). Then, we try to predict what will please—or at least avoid offending them—and do that.²

Experience, however, soon teaches us to abandon the “impossible and absurd project of gaining everybody’s good will and approbation” (TMS, III.2.6.32, 152, n22). First, we learn that people often judge us for things we have not done, or for “motives that had no influence on [our] action” (TMS, III.2.32, 150). Second, through acting to help one person, we will frequently find that we have angered another. If we question why, we will frequently discover that we are judged, not by whether our action was appropriate to our situation, but by whether it was suitable to their interests (TMS, III.2.6.32, 152, n22).

With this double realization, we begin to second-guess people’s judgments of us. We do so by imagining how someone who was not hampered by their lack of information or biased by their vested interests would react to our actions. We begin in other words to imagine how an “impartial and well-informed spectator” would judge us, if such a spectator were present (TMS, III.2.6.32, 152, n22).

Because this imagined spectator is better—“better” informed, “more” impartial—than most real ones, the hypothetical reactions of this ideal spectator gradually take on a normative priority over the actual reactions of most real spectators. The way we think an ideal spectator would react thus becomes how we think a real spectator should react (TMS, III.2.25, 147).

From this point onwards, we always have recourse from the “tribunal” of other people to the “higher tribunal” of “the man within the breast” (TMS, III.2.32, 150): the well-informed and impartial spectator that we can imagine judging us as we ought to be judged.

This is Smith’s story of how we develop a standard of propriety, a standard embodied in the idea of the impartial spectator. For present purposes, I will treat this story as an instance of what Dugald Stewart called “conjectural history”: A history not of how things in fact come about, but

---

² Smith argued that blame and negative emotions hit us harder and are more powerful motivators than is praise and positive emotions (TMS, III.2.15, 141–142). For a more recent treatment of what is now known as ‘negativity bias’, see Rozin and Royzman (2001).
of how they plausibly could have come about, given what we take to be the “principles of human nature” (Stewart 1861, xxxv).3

One of the motivations for offering such a conjectural history is that it can help us identify the main functions of the thing thus explained, or the psychological needs to which it responds. Smith’s conjectural history of the impartial spectator suggests that one of the original functions of this hypothetical spectator is to give us an anchor of stability in the changing sea of others’ opinions of us. By developing an internal moral standard, we can fall back on the “man within” to correct “false judgement[s]” passed on us by ill-informed or partial others, effectively reassuring ourselves that the praise or blame in question is undeserved, and therefore not deserving of our attention (TMS, III.2.32, 150).

This is only one of several functions that a moral standard might serve. But seeing the impartial spectator as an anchor allows us to recognize two problems that may afflict our efforts to employ this standard in everyday deliberation and action: It sometimes gets stuck when we really should be heaving anchor, and sometimes slips when we need mooring the most.4

The second of these two problems, the problem of relying too little on our own conscience, certainly merits the attention it has got in research on how persuasion, conformity, and situation (instigated by Milgram 1963; Asch 1956; and Harman 1999, respectively) affect our moral judgments and actions. Smith himself noted how outside influence may sometimes overwhelm our own considered judgements, leaving us with a warped sense of propriety:

The violence and loudness, with which blame is sometimes poured out upon us, seems to stupify and benumb our natural sense of praise-worthiness and blame-worthiness; and the judgments of the

---

3 As for whether Smith is right about the desire to please and aversion to offend being fundamental parts of our psychology, there is evidence to suggest that autism may be the result of reduced social motivation (Chevallier et al. 2012). If so, a high level of social motivation cannot be constitutive of being human but may be typical.

4 Note that this way of framing the problem neither presupposes nor prejudices the choice of any particular metaethical account of moral truth or absence thereof. To perpetuate the metaphor of the moral anchor: our anchor may always be dragging around in the silt (non-realism), or also sometimes chance upon solid bedrock (realism). The question with which I am concerned is not which of these two descriptions of the seabed is true. I am concerned with the problem, epistemic if you like, confronted by the individual in deciding when to rely on one’s present anchorage (whatever one’s beliefs about the constitution of the seabed), and when to seek new moorings. The metaethical theory you prefer can have an effect on this but need not. This caveat extends to my usage of ‘realist’ and ‘relativist’ in the penultimate section of the article.
man within, though not, perhaps, absolutely altered or perverted, are, however, so much shaken in the steadiness and firmness of their decision, that their natural effect, in securing the tranquillity of the mind, is frequently in a great measure destroyed. (TMS, III.2.32, 150–151)

However, when it comes to disagreements and conflict resolution, the more pressing problem is the first one, namely our tendency to rely too much on our own judgement. Research on the psychology of conflict and disagreement suggests that a significant chunk of our difficulties in this area stems from a kind of inflexibility in our perception and reasoning about other people. This inflexibility gets in the way of giving ground in situations where giving ground would, in fact, be requisite (and appropriate) for resolving disagreement. This inflexibility can be traced to two interrelated biases: The so-called ‘naïve realism’ bias and the ‘fundamental attribution’ error.

II. THE BIASES THAT TRIP US UP

The ‘naïve realism’ bias gets its name from the idea that most people operate on the basic assumption that their idiosyncratic construal of the world is a direct and objective perception of a single, uncomplicated reality (Ichheiser 1949b, 6; Ross and Ward 1995, 278–284).

As first documented in the classic study ‘They Saw a Game’ (Hastorf and Cantril 1954), two groups of people can perceive the same objective reality quite differently, depending on what they are motivated to see. In that study, a student’s home team was consistently seen as more fair than the opposing team (for a more recent example and overview of research on this effect, see Madrigal and Chen 2008).

The effects of motivation on perception have since been demonstrated in other domains, such as politics (see LaMarre, Landerville, and Beam 2009; Kahan, Hoffman, Braman, and Evans 2012), and may run all the way down into our preconscious visual processing (Balcetis and Dunning 2006). Small wonder, then, if we are often unaware of just how our motivations shape our perception of a situation.5

5 The discussion of construal and motivational influences on perception in psychology at some point bleeds into a related but distinct debate about ‘cognitive penetrability of perception’. In that debate, ‘naïve realism’ is the name of a particular philosophical theory, rather than a bias. For more on this debate, see Montemayor and Haladjian (2017), for example.
An apparent corollary of our tendency to naïve realism is our tendency to assume that other people who, like us, are normal and impartial, will see things the same way we do (Ross and Ward 1995, 278; for a recent test of this, see Schwalbe, Cohen, and Ross 2020). To the extent that we are naïve realists, our first impulse when confronted with someone who does not share our opinion may be charitable: They only disagree because they do not know what we know (Ross and Ward 1995, 280). If so, we will also tend to think that it will be easy to get them around to our point of view: We just have to inform them of what we know; present them with the evidence (Ross and Ward 1995, 280).

Often, however, our ‘evidence’ will fail to persuade those who disagree with us. What is obvious to us turns out not to be so obvious to them, and what then? What we do not usually do—conspicuously so—is to take our failure to persuade others as a clue that we might be wrong. Instead, we tend to insist that this is ‘one damn time’ we know we are right and conclude that there must be something wrong with them to stop them from seeing straight (Ichheiser 1949a, 39; Ross and Ward 1995, 278). This is where the ‘fundamental attribution’ error comes into play.

The ‘fundamental attribution’ error is the name given by Ross to the tendency of people (at least in western cultures, see Miller 1984; Henrich et al. 2022) to “underestimate the impact of situational factors and to overestimate the role of dispositional factors” when attributing someone’s behavior to some cause or set of causes (Ross 1977, 183, cf. Sabini, Siepmann, and Stein 2001).6

The logic behind this discrepancy is as follows: if we think we perceive the world as it is, and we think that our reaction is an appropriate reaction, then it follows that our reaction is an appropriate reaction to the world as it is (Pronin, Gilovich, and Ross 2004, 794). If someone else comes along and reacts in a manner that contradicts our own, it follows that their reaction is not an appropriate reaction to the world as it is. The explanation for why they act as they do must lie in the person herself (Ross and Ward 1995, 280–281).7

Let us pause here for a moment to consider how strikingly this picture of humans under the influence of bias fits with that sketched by Smith in

---

6 Fundamental in the sense of being the “first [...] and most frequently cited bias or error” in this area, not in the sense of its place in a hierarchy of biases (Ross 1977, 183). Also called the ‘correspondence bias’ (Gilbert and Malone 1995). For more nuance, see Malle (2006).
7 A recent meta-analysis indicates that this effect is primarily associated with negative events (Malle 2006).
his conjectural history of the impartial spectator: once we come to trust our own judgements, presumed to be reflected through the perspective of an impartial spectator (naïve realism), we begin to see judgements by others that fail to coincide with our own as evidence of their failure to live up to this ideal (fundamental attribution error). If someone else reacts to a situation in a manner we cannot understand, we do not stop to consider whether their construal of the situation might differ from our own. Instead, we think that the other is either ill-informed or partial (or both). And if we are particularly frustrated by our inability to understand the other, we might agree with Frank Wheeler: ‘You’re sick!’

The problem with this is not that other people are never wrong. As Ward and Ross point out, we will often be at least partly right to think badly of those who disagree with us:

Other people in general, and adversaries in particular (so the naïve realist readily observes), rarely hold views or advocate propositions whose acceptance would threaten their economic, social, or psychological well-being. In fact, they generally seem to hold views whose acceptance would plainly advance their individual or collective interests. (Ross and Ward 1995, 281)

But as they go on to note:

What generally will be lacking [...] is recognition on the part of the naïve realist that his or her own interests, ideological beliefs, and construals of facts and evidence are similarly congruent. (Ross and Ward 1995, 281)

In other words, the trouble with naïve realism and the fundamental attribution error is not (primarily) that we are mistaken in thinking that others are partial—they frequently are. The trouble is that we fail to consider that the same may be true of ourselves—and that it frequently will be.

This is part of a broader point: The biases that shape our reasoning are not necessarily bad. Indeed, the case has been made that biases like the ones we discuss here are pieces of a bigger puzzle. Studied in isolation, they may appear ill-suited if not inimical to our ideas of rationality. Once the puzzle is laid, however, a portrait emerges of a person who is pretty close to optimally rational faced with the limited evidence of uncertain quality to which the messy world presents us (Hahn and Harris 2014).
Stereotyping and confirmation bias, to take two other well-known biases that are relevant to moral judgement, may in fact and in the main be adaptive, whether by shaping our beliefs to match reality or the other way around (Peters 2020). The problem is that these biases create paths of least resistance, paths which may lead us astray in certain kinds of situations. One of the most pernicious paths runs from observing differences of opinion to concluding that those who disagree with us are mentally incapable of making correct, impartial judgements.

Seen this way, it would be tempting to conclude that the real issue is a kind of cognitive laziness on our part, and that we should simply try harder to be truly impartial spectators. ‘Try harder’ would then mean something like investing more effort into closing the gap between our take on the reactions of an impartial spectator and the reactions an actual impartial spectator would have, if she were present (TMS III.4.1, 182).

The literature on ‘motivated reasoning’ gives limited reason to be optimistic about our prospects for improving our reasoning through added effort.

III. The Promise and Perils of Motivation

‘Motivated reasoning’ (Kunda 1990) might sound bad by definition, but the motivation in question may simply be the wish to arrive at the right conclusion, whatever this might be. In that case, our reasoning is motivated in a way most people would find not only unproblematic, but admirable.

In the literature, this kind of motivation is usually known as an ‘accuracy goal’ (Kunda 1990, 481), and is sometimes associated with the (self-) image of the reasoner as an ‘intuitive scientist’ (Baumeister and Newman 1994, 4). Simply put, accuracy goals can help us reason better by motivating us to put extra effort into forming our judgement (Kunda 1990, 481–482). If for example you exhort students to be fair and open-minded when mediating a conflict between two college dorm roommates (McPherson Frantz and Janoff-Bulman 2000, Study 2), they will invest extra effort in seeing all sides of the case, evaluating merits and demerits. Indeed, they may even consciously control for what they suspect are their own implicit biases, giving some credence to the idea that trying harder is the solution we seek (Baumeister and Newman 1994, 6–11; Kunda 1990, 481–482).

Unfortunately, an accuracy goal only really works in the absence of its troublesome twin, the ‘directional goal’. If an accuracy goal is the desire to find the right conclusion, regardless of what it might be, a directional
goal is the desire to reach a *particular* conclusion, regardless of what the ‘right’ one might be (Kunda 1990, 482). Operating with a directional goal, the reasoner behaves as an ‘intuitive lawyer’ building a case for a client belief (Baumeister and Newman 1994, 5).

To continue with the example of the college dorm disagreement: you can give the student a directional goal by presenting one of the roommates as gregarious and likeable and the other as overly serious and aloof. This will tend to give the mediator a directional goal to rule in the favor of the likeable student in the dispute between the two (McPherson Frantz and Janoff-Bulman 2000, Study 2). The question now is whether exhorting the mediator to be fair and open-minded can help counteract the biasing effects of this directional goal.

In fact, the opposite happens. The mediator does invest extra effort, but that extra effort goes into building a *stronger* case for the desired conclusion, namely that the more likeable student is right (Baumeister and Newman 1994, 7–8; McPherson Frantz and Janoff-Bulman 2000, 38–41). Having a directional goal tends to make the mediator partial, and adding an accuracy goal on top only exacerbates their partiality. Indeed, this particular combination could be said to produce the *worst* kind of partial judge: one who is supremely confident that she is, in fact, being fair and impartial, thanks to all the effort she has put into reaching her conclusion (McPherson Frantz and Janoff-Bulman 2000, 40). This point is worth repeating: if I am already partial, my partial judgements may only become *more entrenched* if I try harder to be impartial (McPherson Frantz 2006).

It should be noted that this rule has an exception: if the evidence in the case points unanimously and unambiguously towards a different conclusion than the one we desire, we will tend to concede that the desired conclusion is wrong and that the undesired conclusion is right (Baumeister and Newman 1994, 16; Kunda 1990, 487, 490). In the social realm, however, unanimous and unambiguous evidence is hard to come by (Uhlmann and Cohen 2007, 215). As long as we can sift the extant evidence for an argument to ‘persuade a dispassionate observer’ (Kunda 1990, 483), any added effort will tend to go into strengthening the case for the desired conclusion.

Moreover, in contrast to a real defense lawyer, we are not necessarily aware of the fact that our reasoning is under the influence of a directional goal. Indeed, most directional reasoning probably involves a degree of
self-deception, since, under a constraint of rationality, we must also convince ourselves that our argument is sound (Baumeister and Newman 1994, 5; cf., Rosenzweig 2016, on how we may be willing to dispense even with this semblance of consistency). Therefore, the self-image of the directionally motivated reasoner is not typically that of an ‘intuitive lawyer’. Instead, he will tend to think that he is behaving as an unbiased scientist, merely out to get the facts straight (Baumeister and Newman 1994, 5).

In this way, our motivations can play into and exacerbate our naïve realism: we assume that we see the world in a direct and uncomplicated matter, and if motivated to justify our views, we will tend to expend that extra effort on building the case that we were right all along. The way we see things is, in fact, the way things are: “what you see is all there is” (Kahneman 2013, ch. 7).

If you and I both reason in this manner, the effect is liable to be reinforced. Instead of trying to understand each other’s points of view and impartially evaluate the merits of each other’s arguments, we will expend our efforts in bolstering our own sense that we are right. We may allow ourselves to become more and more partial to our own position and willfully ill-informed about that of the other. Research on the role of bias in the escalation of conflict supports this. When one party to a disagreement infers that the other’s lack of comprehension is the result of bias, that party is likely to act more ‘conflictfully’ towards the other. This other in turn perceives (not unjustly) the first party to be really biased, thus acting more conflictfully in return—and so on in a downwards spiral (Kennedy and Pronin 2008).

Thence the painful familiarity of how, in the quote with which I started this discussion, the exchange between April and Frank rapidly degenerates: when Frank calls April “sick” for “twisting” everything he says, April is not thereby persuaded by the superiority of Frank’s (implied sane) perspective. Instead, she replies: “and do you know what you are?” [...] ‘You’re disgusting’, whereupon their quarrel escalates to a fight and goes ‘out of control’” (Yates 2011, 36–37).

In sum, the literature on cognitive biases suggests that we are unaware of the importance of construal to perception and judgement, and all too eager to ascribe differences in opinion to defects in the mind or character of those who disagree with us. The literature on motivated reasoning suggests that we can overcome some of the negative impact of these biases by striving to be impartial spectators, but only if we are not already partial. If we are already partial—or, as is particularly pertinent to the present
discussion, self-partial agents trying to see ourselves from the perspective of an impartial spectator—then aspiring to be impartial and well-informed is likely to result in us being just as partial and ill-informed as ever, with an unshakable confidence that any impartial spectator would agree with our view, to boot.

Admittedly, the picture is bleak. We should, however, not let the negative focus of our review occlude the fact that we sometimes also do quite a good job of seeing ourselves and our social surroundings accurately (Krueger and Funder 2004). Thus, while a simple ‘try harder’ is unlikely to have the desired effect of making us more like our ideal of the impartial and well-informed spectator, this does not signify that we are incorrigible.

Smith, despite all his insights into our frequent failings, also had faith in our ability to do better, and is sometimes quite optimistic on our behalf (see, for example, his discussion about the limits to self-love in TMS, III.3.4–5, 157–159). Charles Griswold suggests we read Smith’s optimistic descriptions of our capacities for sympathy and impartiality as part of a protreptic rhetoric meant to inspire us, his readers, to live up to our own ideals (Griswold 1998, 49, 104). In the context of our present attempt to tease out some better advice than ‘try harder’, this rhetoric merits a closer look.

Smith’s protreptic rhetoric can be seen to comprise two steps. The first step is to show or remind us what our ideal is. Smith takes us through this step in his conjectural history of the advent of the impartial spectator, as related above. The second step is to show us when, how, and why we fall short of this ideal. Smith undertakes this second step over the course of a number of passages that are alternately entertaining and painfully recognizable in their accurate descriptions of our wonderful powers of self-deceit. One example is Smith’s description of the lengths we are willing to go to maintain a positive self-image:

We endeavor by artifice to awaken our old hatreds, and irritate afresh our almost forgotten resentments: we even exert ourselves for this miserable purpose, and thus persevere in injustice, merely because we once were unjust, and because we are ashamed and afraid to see that we were so. (TMS, III.4.4, 183–184)

---

*I do not mean to imply that Smith passes through these steps, only that we can think of the protreptic rhetoric as comprising these two, distinct kinds of revelations.*
After having followed Smith through the two steps of the protreptic rhetoric, we should know what our goal is, and the dangers that we need to avoid in order to attain that goal. The psychological research I have reviewed so far complements the second step of the Smiths' protreptic rhetoric by revealing in detail when, how, and why we make mistakes in our judgements of others and become overconfident in the probity of our own.

But how, exactly, does going through this second step equip us for the struggle to be more like the impartial spectator identified in the first step? The most obvious answer would be that, by being made aware of the operation of such biases, we are now in a position to overcome the negative impact they have on our judgements (for evidence supporting this approach, see Petty et al. 2007, 269–273; Nasie et al. 2014).

This, however, would be premature. The problem is not just that we are unaware of the impact of bias on our judgements: people frequently continue to believe themselves to be free of bias after being made aware of its likely impact on them, and even as they condemn others for failing to take biases into account in their judgements (Hansen et al. 2014; McPherson Frantz 2006; Pronin, Gilovich, and Ross 2004, 788; Pronin, Lin, and Ross 2002). On its own, therefore, mere awareness of bias is unlikely to be an effective remedy (Bezrukova et al. 2016). For this knowledge to be turned into a power to do better, we must find some way of integrating it into our striving.

Consider the following: the problem pointed out by research on naïve realism is that we tend to rely too much on what we take to be the perspective of an impartial spectator. This creates a world of trouble because what we take to be the perspective of an impartial spectator is invariably a version of our own perspective, which will often, if not always, be less than fully informed and impartial. That said, we can do better if we try: the potential for accuracy goals to improve our reasoning partly vindicates this premise of Smith's protreptic rhetoric.

What matters, then, is how we try. Simply trying harder will not do: given that accuracy goals exacerbate rather than ameliorate the biased reasoning associated with directional goals, trying harder to be impartial spectators is apt to backfire, making us even less like the impartial spectators we strive to be. The answer thus cannot be 'try harder', it must be 'try smarter'.

IV. TRYING SMARTER
As for how we should ‘try smarter’, it is worth remembering that the whole process leading up to the idea of a hypothetical impartial spectator starts out with our imagining the perspectives of actual spectators in an effort to predict what they will like and dislike. Only because of inconsistencies in the approval and disapproval of these actual spectators do we gradually shift towards imagining the reactions a hypothetical ideal spectator would have in their stead.

Seen from this perspective, the evidence from research on the impact of bias and motivated reasoning on interpersonal judgement and conflict reveals that many of us shift too far towards taking only the perspective of (our idiosyncratic construal of) the ideal spectator. An element of humility appears to have gone missing: we assume that we are right, we assume that any impartial spectator would agree that we are right, and we assume that anyone who disagrees with us does so out of a failure to take the perspective of such an impartial spectator. If this is indeed our problem, there would seem to be a simple solution: being a bit more humble.

But how, you may ask, are we to re-introduce the element of humility that appears to have gone missing (if it was ever present in the first place) into the reasonable aspiration to be more like the impartial spectator? Consider this: if we have shifted too far towards only taking the perspective of the hypothetical impartial spectator, then we have also shifted too far away from taking the perspectives of the real spectators around us. Trying smarter must therefore include shifting the balance back towards taking the perspective of other people.

Intriguingly, there is evidence to suggest that naïve realism and other psychological barriers to conflict resolution can, in fact, be addressed by getting the opposing parties to take each other’s perspective (Batson and Ahmad 2009; Baumeister and Newman 1994, 7; Bruneau and Saxe 2012; Galinsky and Moskowitz 2000; Long and Andrews 1991). It must, however, be done properly. Overcoming the negative impact of naïve realism requires more than just ‘activating’ our ability to perspective-take, although this is an important first step (Epley and Caruso 2014, 300). Activation is insufficient because in trying to adjust our perspective to bring it closer to that of someone else, we may adjust too little, too much, or in the wrong direction, and yet, in an echo of motivated reasoning, think that we are doing it right (Epley and Caruso 2014, 302–307).
Therefore, merely reminding ourselves to imagine the perspective of the other will not be enough. Indeed, those who are the worst at understanding other people also think they are the best at it (an example of the Dunning-Kruger effect, see Ames and Kammrath 2004). Therefore, those who are most likely to fail to properly consider the point of view of the other are also those who are most likely to think that they are already taking it into consideration. Trying smarter cannot, therefore, merely be a matter of reminding ourselves about the importance of perspective-taking. Doing so may just make us that much more indignant in the failure of others to take ours.

Nor can trying smarter, as we have already seen, be a matter of exhorting ourselves to be ‘impartial’ or ‘unbiased’. This will often serve no other function than to entrench our biased beliefs. Nor, finally, can it be a matter of informing ourselves about the negative impact of bias. This, as we have also seen, will frequently do little more than confirm us in our belief that everybody else is biased beyond redemption, and hence unworthy even of the effort to see things from their (‘sick’) perspective.

We need to increase both the likelihood that we will consider the point of view of the other, and the quality of our perspective-taking. How do we do that? There is really only one way of improving the quality of our perspective-taking. We must let ourselves be informed by “the person principally concerned” (TMS I.i.4.6, 26). We must listen to the other.\(^9\) Since this will not happen by itself, we need some way of reminding ourselves to listen to them. Especially when we are least disposed to do so. What can serve this function? I would like to argue that the ideal of the impartial spectator can serve this function—if we reshape it.

Consider again the second step of Smith’s protreptic rhetoric. It shows us when, how, and why we make mistakes in our judgements of others and become overconfident in the probity of our own. This knowledge is meant to help us be more like impartial spectators. Above, we saw that knowing about biases is not, on its own, enough to make us more like impartial spectators. The second step of the protreptic rhetoric thus appears to fail its purpose. But we can draw a different conclusion from these results if we look at the two steps of the rhetoric under one:

\(^9\) Indeed, in a study of people’s tendency to give too little weight to the opinions of others in making accurate judgements, Liberman et al. (2012) found that this ‘underweighting’ phenomenon was mediated by naive realism, and that this could best be eliminated by asking people to talk their way to a shared estimate.
If we strive for the ideal of the impartial spectator, then knowing about directional goals and naïve realism will not in itself help us and may, in fact, only make matters worse.

The failure of the second step thus invites us to reconsider the first step in light of this failing. We should, that is, consider what role the ideal of the impartial spectator itself plays in generating the kind of trouble we find ourselves in, and whether there is something to be done to minimize this negative impact while maintaining the positive influence of having an ideal to which to aspire.

V. A PARTIAL SOLUTION
We have already established that ‘trying smarter’ must include shifting the balance away from the impartial spectator, and back towards the perspective of other people. This is right in the sense that we need to invest more effort into understanding others compared to defending ourselves. Battling the beast of bias requires a return to the basics of ‘sympathy’ as Smith understood it, a return to the effort to “bring home to our own breast” other people’s situation (TMS, II.i.5.3, 88).

We cannot, however, abandon the ideal of the impartial spectator only to go chasing after other people’s incongruous ideas of praise- and blame-worthiness. We have already been there, tried that, and rejected it as an unsupportable mode of living; doing so would be a reiteration of the “impossible and absurd project of gaining everybody’s good will and approbation” (TMS, III.2.6.32, n22, 152) that led to the invention of the impartial spectator in the first place. Abandoning the ideal would just restart the cycle.

The solution, therefore, is not to abandon the ideal, but to reshape it.10 We must reshape the ideal spectator because the first casting produced an ideal that was simultaneously too malleable and too inflexible: Too malleable because too easy to mold into a model of ourselves, so that considering what an impartial spectator would think too easily turned

---

10 This reshaping of the ideal of the impartial and well-informed spectator also serves a response to the claim, in Forman-Barzilai (2010, 159), that Smith’s ideal is practically incoherent because of a deep tension between being impartial and being well-informed. As Sivertsen (2019, 66) argues, the ideal of the impartial and well-informed spectator may yet be rescued as an ideal for practical moral deliberation if it is reshaped to be less ideal, more psychologically realistic. The partially impartial spectator is my proposal for such a psychologically informed, non-ideal, regulatory conception of the ethical ideal of the impartial spectator.
into an exercise in convincing ourselves that we were impartial all along; too inflexible because too difficult to change once set in this mold.

Note, however, that my reshaping the ideal is undertaken for the express purpose of improving our chances of actually being impartial spectators. The theoretical, ethical ideal of the impartial spectator is to be kept, but it must be kept thoroughly in the background, at least in the kind of interpersonal dispute that we are dealing with here. What we are after, therefore, is a new foreground or regulative ideal, something for which we can aspire and through which we can actually improve. To borrow a pair of well-known terms from political philosophy: we need a non-ideal principle to guide us, one that takes into consideration the biases and limitations in our thinking, in order to bring us closer to the ideal principle of the impartial and well-informed spectator.

The main task of this foreground ideal is to increase the likelihood and quality of our perspective-taking with the other(s) involved in the dispute. What shape must we give our mold to achieve this? It must, I think, be what I will call a sophisticated relativist.11

The sophisticated relativist is different from the naïve realist in two ways: she asserts i) that many, if not all, social situations of moral import—like that of an interpersonal disagreement between friends or lovers—admit of several, apparently incompatible, valid construals (that’s the relativist bit), and ii) that her construal might not be one of the valid ones (that’s the sophistication).

You might wonder whether it would not be enough for our foreground ideal only to assert ii), leaving open the question whether there is a single truth of the matter in moral matters—or even positively asserting that there is a single valid construal of most if not all social situations of moral import. You might wonder, that is, whether the smallest change needed to shape the impartial spectator into a usable foreground ideal is not simply to make it a sophisticated realist. After all, if our problem is our naïve belief that our construal is the only right one, we only need the humble admission that it might not be to get us closer to true impartial spectatorship.

It might be true that all we need is ii), but, as I have already argued at length, the kind of self-critical thinking that is needed to acknowledge ii) is hard to come by (for further evidence supporting this, see Lilienfeld,

---

11 The ‘relativist’ label is only intended as a rough and ready complement to the ‘realist’ of naïve realism, inspired by Donna Haraway’s discussion of realism and relativism in science (1988).
Ammirati, and Landfield 2009). The whole point of the exercise with which I am presently engaged is to discover some other route to that realization, some route that, if habitually followed, might allow for the development of the courage requisite to go directly to ii)—a courage Smith compares to the boldness of a “surgeon … whose hand does not tremble when he performs an operation on himself” (TMS, III.4.4, 183). This is why we need the assertion that many, if not all, situations of moral import admit of several valid construals.  

By ‘we’ I mean people like me: partisans to the cause of I, who nonetheless aspire to be impartial spectators. We suffer from the naïve realism bias and are all too apt to fall into the trap of the fundamental attribution error when confronted with someone who does not share our view. We need a foreground ideal to remind us that we are not the only ones in the room who might be right, and the surest route, I claim, goes through i), the relativist aspect of the sophisticated relativist.

Note, however, that the sophisticated relativist, in contrast to what—following the naming convention I have adopted here—may be called the naïve relativist, does not accept all takes on a situation as equally valid. The sophisticated relativist knows that the real problem is naïveté: merely substituting relativist for realist notions may shift the battleground, but it does nothing to resolve the conflict.  

Hence, the sophisticated relativist insists that our construal and judgement are only worthy of consideration if they are the result of an honest attempt at being impartial. The-impartial-spectator-as-sophisticated-relativist is still an aspirational ideal, still something we have to strive for.

Aspiring to be a sophisticated relativist thus means striving to see oneself and others as one imagines an impartial spectator would see us, while always acknowledging that whatever impartiality we may lay claim to is at best a partial impartiality. Partial not in the oxymoronic sense, but

---

12 Another option, worthy of further study, is the potential application here of the Jainist principle of anekantavada, or ‘no-one-perspectivism’, often illustrated by the fable of the five blind men and the elephant. Each of the men, perceiving only an isolated part of the whole, thinks that it must be a tree (the trunk), a rope (the tail), a wall (the side), a pillar (a leg) or a huge fan (an ear). A sixth man resolves the dispute by explaining how each of them is partially right because they each perceive a part of the whole elephant. Hence, a complete understanding of the truth of any matter (the satya) requires the consideration and acceptance of multiple, apparently incompatible points of view (see Mehta 2018).

13 In the case of naïve realism, we admit only the truth of our own point of view. In the case of naïve relativism, we admit only the truth that all points of view are equally valid. In both instances, we fail to admit the possibility of error in our own point of view and truth in the perspective of another (see Haraway 1988, 584).
in analogy to Donna Haraway’s ‘partial objectivity’ (1988): more demanding and more impartial than the naïve embrace of one’s idiosyncratic perspective, but still only a part of a bigger picture, a picture others who are likewise engaged in striving to be *partially impartial spectators* must also contribute (see also Freiin von Villiez 2006).

This invites the question of how we are to identify those others who are thus engaged. How are we, that is, to identify those whose diverging opinion we have to take seriously? How do we decide when to listen and give ground, and when to speak and stand fast? This is a difficult question, and I am not sure that there is an easy answer. There is, however, a method we can employ that may help us discover whether our interlocutor is interested in resolving the dispute through mutual rapprochement.

This method is simply to *unilaterally* apply the ideal of the partially impartial spectator in a dispute (see ‘GRIT’ in Osgood 1959). In other words, instead of waiting to discover whether our opponents are willing to listen to us, we should start by listening to them. If our opponents are similarly motivated, they will eventually reciprocate. If they do not reciprocate, we will have an indication that they are not motivated to reach an understanding, and we will have to decide whether to continue with our effort to resolve the dispute.

In short, having the partially impartial spectator as a moral ideal means always being ready to listen to the other because regardless of how certain we feel that we possess the truth, this is at best only a part of a bigger truth, and our opponents may well, regardless of how certain we feel that they do not, possess other parts of that truth.

Another worry remains. What if the only effect of adding ‘partially’ to ‘impartial spectator’ is to replace the conviction that we are right with the conviction that we are right enough?

What if, that is, the effort we are trying to channel away from the construction of ever more elaborate rational justifications for fundamentally biased beliefs does not end up going towards the amelioration of that bias through better perspective-taking? After all, these energies may be turned to other projects, or simply dissipate.

This is indeed a real danger, and one I cannot see any easy way of heading off. One imperfect option is to point to the fundamentally social nature of our being and reiterate Smith’s point about our desire for mutual sympathy. This is an imperfect option, because we may seek that community where there is least resistance to our views. It is nevertheless
a worthwhile option to consider, since the experience of mutual sympathy, the experience of being able to reach an understanding with someone, is inherently positive, and thus may serve to reinforce our commitment to trying smarter in being impartial spectators.

In the end, my whole argument amounts to a wager. I wager that holding something like the partially impartial spectator as a moral ideal will improve our chances of resolving the kind of interpersonal dispute exemplified by April and Frank Wheeler in the opening quote of this article.

Whether this wager can ever be fully decided by empirical research is, I think, an open question. We might, however, get some indication of whether the ideal of the partially impartial spectator has any merit by plugging a description of this ideal into established empirical paradigms for testing, for example, the effects of accuracy goals on directional reasoning (e.g., McPherson Frantz 2006). If thinking like a partially impartial spectator ameliorates the negative impact of bias in such cases, this would be an indication that I am right.14 If the opposite turns out to be the case, and the ideal of the partially impartial spectator exacerbates bias as much as, or more, than do exhortations to be ‘fair’ and ‘unbiased’, this would be an indication that I am wrong.15

VI. CONCLUSION
The naïve assumption that we see the world as it is, in a direct and uncomplicated manner, gets in the way of our seeing it impartially. It does so because instead of listening to those who disagree with us, we too easily conclude that there must be something wrong with them to stop them from seeing what to us is plain truth.

Knowing about this unfortunate tendency of ours is not enough, on its own, to bend our deliberations in a better direction. As long as we are trying to be impartial spectators, we will continue to fall into the trap of mistaking our attempts at being impartial for actually succeeding, and in this mistake, condemn those who disagree with us as hopelessly partial.

What we need, therefore, is to redirect our strivings away from that ideal, and towards other people. Only by being less concerned with what a hypothetical impartial spectator would think of our disagreement, and

---

14 For evidence of the converse effect, see Uhlmann and Cohen (2007). Also, for overconfidence, see Harvey (1997), and for a possibly analogous case of the impact on judgement of different explanations of an apparently self-evident concept, see the case of ‘reasonable doubt’ in Horowitz and Kirkpatrick (1996).
15 For what it’s worth (which is not much), my own, very personal experiments with this are so far inconclusive.
more concerned with what actual others think, can we hope to balance our belief in our own pre-eminence with willingness to listen to others who might see us less favorably.

However, we cannot simply abandon our striving to be more like the impartial spectator. Even if we are, at present, egregious egocentrists, our problems are not solved by becoming abject allocentrists. Therefore, we need a reshaped ideal for which we can strive, and in striving for which we can actually improve.

In practice, a partial impartiality is the best any of us can hope to achieve on our own. Incorporating this limitation in our personal conception of the ideal spectator, we combine intellectual honesty with ethical humility and openness to others into the very thing for which we strive. If we thus choose the partially impartial spectator as a regulative ideal of our own moral judgements, then no matter how great our capacity for self-delusion, we will always, if we stop to consider it, recognize the potential for improvement in what are necessarily imperfect moral judgements, and be reminded of the need to listen to those who disagree with us.

When Frank Wheeler jumps from his certainty that he is right—“This just happens to be one damn time I know I’m not in the wrong”—to the conclusion that there must be something wrong with his wife April—“You know what you are when you’re like this? […] You’re sick. I really mean that”—he is operating on the default logic of naïve realism. If I am right, and I think any impartial spectator would agree, then, if you refuse to admit that you are wrong, there must be something wrong with you. While it may already be too late for April and Frank at this point, breaking the logic of naïve realism is one way of stopping the downwards spiral of disagreement into conflict.

Adopting the partially impartial spectator as a foreground ideal for moral deliberations breaks the logic of naïve realism by breaking the link between my certainty that I am right and my belief that you must be wrong. It does so by replacing, in practice, the monolithic ideal of the impartial spectator with a multifarious one. I may still believe that I am right, and I may be convinced that I am a partially impartial spectator of myself, but this does not mean that there is anything wrong with you, since two partially impartial spectators may well differ in their opinions without either being ‘sick’ or ‘disgusting’. Or so both of us might think, and so both of us may be that more willing to listen to the other.
Whether this will work or not is, at least in part, an empirical question, and given the potential benefits, I think it worth investigating further.

REFERENCES


Stewart, Dugald. 1861. “Account of the life and writings of Adam Smith.” In The theory of moral sentiments; or, an essay towards an analysis of the principles by which men naturally judge concerning the conduct and character, first of their neighbors, and afterwards of themselves, to which is added a dissertation on the origin of languages (new ed.), xi–lxix. London: Henry G. Bohn.

Sveinung Sundfør Sivertsen is a guest researcher at the University of Bergen and a philosopher at the Norwegian NGO Fremtenkt. Contact e-mail: <sveinung.sivertsen@uib.no>
Anodyne Privatization

JOSEPH HEATH
University of Toronto

Abstract: Privatization of state services has been a flashpoint for political conflict over the past several decades. The goal of this paper is to explain why someone who is a supporter of the welfare state might also support the privatization of certain state services, in certain cases. Recent philosophical literature has focused on the most problematic privatization initiatives, especially the introduction of private prisons and military contractors. As a counterpoint, this paper describes a set of anodyne privatizations, understood as privatizations that no reasonable person could object to. The key step in this analysis involves showing that privatization is not a unitary phenomenon. There are different types of privatization, different degrees of privatization, and also different motives for privatization. There are also important normative differences between these initiatives, which might lead a reasonable person to support some but not others.

Keywords: privatization, welfare state, public sector, state-owned enterprises, outsourcing

JEL Classification: H4, H6, L3, P16

I. INTRODUCTION
Public debate over the privatization of state services often exhibits a Manichean tone, as though the boundary between the public and private sectors was the site of an epochal struggle between good and evil, with the state on one side, representing the forces of justice, and the private sector on the other, representing the greed of capital. Discussion of the issue among political philosophers has been somewhat more nuanced, yet there is still a tendency to treat privatization as a unitary phenomenon, the desirability of which can be determined in the abstract, on the basis of fundamental principles of justice.¹ On top of this, the assumption

¹ See papers collected in Dorfman and Harel (2021). For a representative instance, see Cordelli (2020) and Herzog (2021). Also Knight and Schwartzberg (2019).
remains widespread that all those who support the welfare state, along with its central mission of attenuating the injustices generated by the market, should be opposed to privatization in all forms. Thus a great deal of energy has been invested in the task of producing increasingly clever arguments against privatization (with particular emphasis on considerations that rule it out on deontological grounds), without much attention to the administrative difficulties that arose within the public sector that provided the primary motivation for it (Dorfman and Harel 2013, 2016; Cordelli 2021).

My central ambition in this paper will be to explain why someone who is a supporter of the welfare state might also support the privatization of certain state services, in certain cases. A great deal of the philosophical literature has focused on the most problematic privatization initiatives, especially the introduction of private prisons and military contractors (DiLulio 1988; Pattison 2010). As a counterpoint to this, I would like to describe a set of anodyne privatizations, understood as privatizations that no reasonable person could object to. This provides a useful framework for assessing the acceptability of any particular privatization proposal, because most can be situated between these two extremes. The first step in developing this analysis involves showing that privatization is not a unitary phenomenon. Most importantly, there are different types of privatization, and confusingly, different degrees of privatization. There are also importantly different reasons for privatization. Once it is made clear that a half-dozen quite different phenomena are routinely lumped together under the heading of ‘privatization’, it becomes far less surprising to discover that there are important normative differences between these various initiatives, which might lead a reasonable person to support some, but not others. Thus the normative assessment of privatization must be a great deal more nuanced than it has been, or than the current tenor of the philosophical literature encourages.

II. Preliminary Clarifications

As a prelude to any detailed discussion of the merits of privatization, it is necessary to issue a few clarification and qualifications. First of all, it is essential to distinguish core functions of the state from what are often described as its ‘economic’ activities (Stiglitz 1989). It is the latter that have most often been the focus of privatization efforts. There are a variety of ways of drawing the core-periphery distinction, some more tendentious than others. Perhaps the least controversial is to begin with the
Weberian definition of the state as the institution that exercises a monopoly on the legitimate use of force in a territory (Weber 2004). The core components of the state can then be defined as those that are involved in constituting the institution with this power. This will include not only the agencies involved in securing and preserving that monopoly of coercive force de facto (that is, the military, police forces, the penal system), but also those involved in establishing the conditions of its legitimate exercise (that is, the legislature and judiciary). Although there has been some refinement of technique, these components of the modern state have been in place for centuries. In more recent years, however, it was discovered that an institution possessing a monopoly of force could also be an important economic actor, and indeed, that it could accomplish certain tasks that no private actor seemed prepared to take on. The result has been a massive expansion of state economic activity, particularly in wealthy industrialized democracies, where state spending as a fraction of GDP typically rose from less than 2% at the beginning of the 20th century to roughly 40% at the end (Lindert 2009). Areas in which the state has become a particularly prominent actor include banking, education, health, infrastructure, mass transportation, communications, and insurance (including pensions) (Moss 2002; Barr 2020).

Most of the political debate over privatization has been focused on these economic functions, since this is where the bulk of state spending is concentrated. As a result of privatization, most Western states have withdrawn from ownership of automobile manufacturing (Jaguar, Renault), airlines (British Airways, Lufthansa, Air Canada), retail banking (BNP Paribas, Credit Lyonnais, Thomas Cook), railroads (British Rail, Canadian National), petroleum (British Petroleum, Elf, PetroCanada), mining (Charbonnages de France, British Coal, PotashCorp), and telecoms (Telstra, France Télécom, Deutsche Bundespost Telekom). There were particularly acrimonious debates in the 1980s and 90s over privatization of the water supply and railroads in the U.K., telecoms in France, electricity systems in the U.S., as well as the state petroleum company in Canada (Donahue 1989). The current philosophical debate, by contrast, has tended to focus on privatization initiatives that impinge upon the core coercive

---

2 It should be noted that the state also uses its coercive power to engage in a certain measure of redistribution, primarily through the tax system and entitlement programs. Privatization is seldom an issue here, because only the state possesses the coercive power to achieve redistribution, and so these tasks cannot be privatized. An exception, to be discussed below, is when the state is engaged in in-kind provision of some good and switches to cash or voucher-based redistribution.
apparatus of the state, such as private prisons or the use of military contractors. Here it would not be surprising to discover that there are strong objections to privatization, stemming from basic philosophical principles.

It is generally agreed that the state does not arise spontaneously, through a mere confluence of individual interests or out of voluntary contractual relations. After a coercive public power has been established, then individuals can accomplish a great deal through private contracting, but these contractual relations are only binding because of the background guarantee offered by the public authority; they cannot serve to establish that power. As a result, the core coercive apparatus of the state operates on a different logic from that of any private institution. It is, therefore, courting disaster to begin replacing elements of that apparatus with private organizations. Most obviously, because of the dangers of corruption, there is considerable merit in compensation systems that neutralize pecuniary concerns for state employees who are directly involved in the exercise of coercive power. It is, of course, important to avoid the fallacy of thinking that those who work in the public sector automatically acquire a commitment to the public interest (or that individuals in the private sector can be expected to act in a self-interested fashion, merely because the organizations in which they work are committed to maximizing profit) (LeGrand 2011, 88). This involves a confusion of individual motivations with organizational objectives. It is, however, important to observe that public-sector organizations are in a position to institute management models that significantly reduce the risk of corruption or abuse of power in a variety of ways: by having strong employment security, relatively dull financial incentives (for example, seniority-based compensation), strong shared values (including patriotism, the rule of law, and equal citizenship), and a culture of deference to elected officials. There is such a thing as a ‘public sector ethos’, particularly at higher ranks of the bureaucracy, even though it stops short of making public officials purely disinterested (Pratchett and Wingfield 1996, 639–656).

In part for these reasons, most advanced welfare states have not been especially keen to privatize core state functions. There is also the simple fact that in most countries these functions are not particularly expensive, and so privatization does not have much upside from a fiscal point of view. Two prominent exceptions are the U.S., which maintains an unusually large prison population of over 2 million persons, with another 4 million on probation or parole (Harding 2001), and Israel, which maintains
what amounts to a permanent military occupation over a subject popula-
tion of more than 3 million (Seidman 2010). It is perhaps not surprising
that, in these countries, the temptation to privatize elements of the coer-
cive state apparatus is much greater than elsewhere, and that the contro-
versies generated have occupied the attention of academics (Medina
2010). In most other countries, however, these issues are not particularly
pressing. And so, in the discussion that follows, I will focus my attention
only on the privatization of economic functions of the state. Apart from
the fact that this is where most privatization has occurred, this is also
where one is likely to find the more uncontroversial instances of it.

The second major qualification worth registering is that, in a complex
mixed economy, the public sector is almost never public ‘all the way
down’. In other words, if one examines a particular public agency, en-
gaged in the provision of some public good or service (for simplicity I will
refer to both as ‘goods’), there will inevitably be some elements that are
‘contracted out’ or purchased from private suppliers.\(^3\) The U.S. Postal Ser-
vice, for example, operates over 300 vehicle maintenance facilities and
employs over 4,000 mechanics, in order to keep its fleet of 150,000 deliv-
ery vehicles in a state of good repair. At the same time, it does not actually
construct the vehicles—the ‘Grumman LLV’ that they use was built by
Grumman and General Motors. This is fairly typical—if one digs deep
enough, one will eventually come upon inputs to government activity that
are supplied by the private sector. This boundary is often easy to discern,
because most states are required to go through a public tendering process
in order to acquire these inputs. The set of rules governing this acquisi-
tion process is extremely complex and forms the cornerstone of the ar-
cane but enormously important domain of public administration known
as ‘procurement’ (Thai 2001).

This observation is philosophically significant because it serves as a
reminder that the distinction between public and private is often not ei-
ther-or, but rather a question of fixing the boundary at which the public
stops and the private begins. Failure to attend to this leaves some argu-
ments against privatization, which impose an overly categorical prohibi-
tion, vulnerable to a reductio, because they wind up prohibiting any sort
of procurement from the private sector. For example, if one appeals to
distributive justice considerations to claim that the state should hire its

\(^3\) This is true even with respect to the coercive apparatus. Towing and impounding of
vehicles, for example, is often performed by private contractors. Bailiffs (used to enforce
payment on civil judgments, including repossessions and evictions) are also private in
many jurisdictions.
own cafeteria workers, rather than contract out food service to a private company, on the grounds that the state pays higher wages, then it follows that the state should not be purchasing semi-prepared frozen foods either, but instead should be doing all food preparation in-house, because the state can pay its prep workers a higher wage, or that the state should not be purchasing food from private farmers, but rather should be operating its own agricultural facilities, because the state can pay its growers and pickers a higher wage, and so on. Anywhere that one attempts to draw the boundary between public and private, there will be an obligation to extend the public further, and so the argument amounts to a requirement of economy-wide socialism.

The other thing worth noting about this issue is that, at a certain level of granularity, the question of where the boundary should fall between public and private seems unlikely to be one that can be settled through appeal to first principles (Stark 2009). In practice, it typically gets made on the basis of simple considerations of cost, quality, and convenience (or ‘net value’) (Moore 1995). In this respect, public managers engage in deliberations that are strikingly similar to those engaged in by private managers confronting the ‘make or buy’ decision, the analysis of which has figured centrally in the transaction-cost theory of the firm. On this view, the boundary of the firm is determined by the scope of the employment relation. Various firm inputs, including labor, can be obtained through market contracting. The employment relation is also contractual, but differs in one key respect, which is that its terms are incomplete. Rather than agreeing to perform a precisely delineated task in return for some consideration, the way that a contractor does, an employee instead agrees to a relatively open-ended job description, but under the authority of management. In other words, the employee agrees to perform tasks as assigned.

The choice between public and private, at the boundary, typically involves the exact same question about the capacity of the state to achieve its objectives through contracting, rather than through its own employees. For example, while Canada Post also operates maintenance facilities for its delivery vehicles (mainly Grumman LLVs), more extensive refurbishment is contracted out to a private firm (Jasper Engines and Transmissions). While there may be some good reason, grounded in basic principles of political philosophy, for the postal service to be in the public sector, it seems rather unlikely that these same principles can resolve the question of how extensive a repair program must be before it makes sense.
to contract it out to a firm that specializes in engine rebuilding. The latter question is often a very difficult one, which involves primarily comparing the state's 'agency costs' (those involved in getting its own employees to do it) to its 'contracting costs' (those involved in getting an outside party to do it). These costs include not only the direct budgetary costs, including both salaries, payments, and the cost of supervision, but also indirect costs, such as delay, inflexibility, and opportunism, as well as hidden costs, such as failures of performance, quality defects, and shifts in risk-bearing.

Public managers are, in my experience, neither stupid nor particularly ideological, and so typically are in a position to recognize when they are able to draw up contracts and oversee them effectively and when they are not. There is a major body of work in public administration—somewhat absurdly named—of ‘deliverology’ dedicated to the question of how to deliver results in government, in which questions about contracting, and the contractual specification of ‘deliverables’, looms large (Barber 2015). Contrary to the narrative of rampant neoliberalism, which registers only a one-way process of selling off state assets, there is actually a great deal of both out-sourcing and in-housing at the boundary of the public sector, along with significant movement back and forth (that is, not just privatization, but also reversals of privatization) (Hefetz and Warner 2004). There is also a great deal of renegotiation of these contracts over time (Perard 2009, 202). So it is important to recognize that a great deal of privatization is not driven by political ideology or rival conceptions of justice, but by pragmatic managerial decision-making that merely adjusts the inevitable boundary between public and private.

III. The Traditional Public-Sector Model

Much of the recent philosophical literature on privatization has not been informed by a particularly fine-grained appreciation of how state operations are carried out, or of how much variation in institutional practice exists from one jurisdiction to the next. In particular, when speaking of the ‘public sector’, many philosophical discussions appear to presuppose what Evan Davis refers to as the ‘traditional public sector model’ (TPSM) as the standard for service delivery (Davis 1998). The key characteristic of this model is that it involves the provision of some good by employees of the state, funded through general taxation, such that it is free to

---

4 For discussion, see Heath (2019).
citizens at the point of service. A paradigmatic example of this is the way that policing is funded and delivered, or the way that primary education has traditionally been provided. The crucial feature of this model is the fact that the state stands on both sides of the transaction—as both the purchaser (that is, the source of funding) and the provider.

One can see the power that this model exercises over the philosophical imagination in the way that debates over public health care, or ‘socialized medicine’ have been conducted. Here the image of the National Health Service (NHS) in the U.K. looms large. What is striking about the NHS is not just that patients are not charged any fees at the point of service, but that doctors and nurses are salaried employees of the state. While granting that there is something to be said for this arrangement, it is important to observe that it is extremely uncommon as a model of socialized medicine. The more typical arrangement, such as one finds in Germany or the Netherlands, is one in which the state intervenes quite heavily in the health insurance sector, in order to guarantee universal coverage, but then leaves a great deal of the provision of health care services to the private sector. Furthermore, doctors are often not salaried, but rather are compensated on a fee-for-service basis. The absence of point-of-service fees (or co-payments) is also not universal—Sweden for instance charges user fees for physician appointments, hospital stays, and emergency room visits (Glenngård and Borg 2019).

More generally, it is extremely common to find arrangements in which the state acts as the sole provider of a good, but where citizens must purchase it in the same way that they would any other commodity. For example, even when the state acted as the sole provider of home electricity, consumption was always metered, and individuals had to pay for the amount that they used. (The old state-owned British Gas even installed coin-operated gas meters in many homes, so that residents would have to keep a supply of shillings on hand, in order to feed the meter when it began to run low.) The state often tacitly subsidized consumption by keeping rates low, but there was never any suggestion that consumption should be entirely funded by general taxation, and thus be free to individuals at the point of service. More generally, while the U.K. Labour Party was once committed to state ownership of the means of production, the goal was never to make the products of companies such as British Steel or British Leyland, much less Rolls-Royce, freely available to all.

---

5 For an overview, see Emanuel (2020).
There are also many cases in which the state acts as the sole (or primary) purchaser of a commodity (that is, a monopsonist), but where supply is completely private. Road construction is perhaps the most common example of this. The state is practically the only purchaser of large-scale roads and highways—which it typically then makes available ‘for free’ at the point of use to all citizens—but the actual construction of these roads is almost always done by private contractors. Construction of government buildings and other infrastructure, such as bridges, is the same, not to mention smaller items, such as signage, garbage cans, tree planters, and public art. These areas all feature a ‘purchaser-provider split’, where the state acts either as the purchaser, or the provider, but not both. The central feature of the traditional public sector model, by contrast, is that it involves the state acting as both purchaser and provider.

Given the ubiquity of purchaser-provider splits in the public sector, one might wonder why the TPSM has occupied so much attention in the theoretical literature. One reason is no doubt the influence of Paul Samuelson’s definition of ‘public goods’ as those that are perfectly non-rival and non-excludable (Samuelson 1954). This has given rise to a great deal of confusion, since practically no goods provided by the public sector possess either quality. If one thinks of insurance, for example, and insurance-like arrangements, which constitute an enormous fraction of the economic activities of the welfare state, not only are the payments perfectly rival, but it is in most cases easy to exclude individuals from the benefits. But nevertheless, for goods that do happen to be relatively diffuse and optimally shared by many, the TPSM typically represents the most efficient mode of provision.

A non-excludable good is just the limit case of a good that generates positive externalities. If the entire value of the good takes the form of an externality, then it may make sense for the state to produce the good and make it freely available to all (think, for example, of direct air capture carbon scrubbing). But if the good generates some mixture of private value to individuals and positive externality for others (as with various forms of green energy), it may make more sense for the state simply to subsidize production of the good, or make it available to individuals at a below-market price, in order to promote optimal production of the externality. In other words, the appropriate price to charge for such a good is dictated by the fraction of the good that constitutes private value (that is, appropriated by the individual), and the fraction that takes the form of a positive externality. Making goods free at the point of service is
appropriate for Samuelson public goods only because these are goods whose value would be impossible to monetize if an attempt were made to provide them privately.

The second major reason for an overemphasis on the TPSM is normative, stemming from the conviction that an important function of the welfare state is to achieve a ‘decommodification’ of certain goods. Michael Walzer is no doubt the most influential philosophical proponent of this view, with his claim that certain goods, such as health care, belong in their own 'sphere of justice', in which they are distributed in accordance with principles that are fundamentally different from those governing the distribution of market commodities. (His discussion echoes the NHS constitution, which states as its second fundamental principle that access to NHS services should be “based on clinical need, not an individual’s ability to pay”—DHSC 2021.) For this reason, Walzer’s vision of a public health care system is one in which “ownership is abolished” and “practitioners are effectively conscripted or, at least, ‘signed up’ in the public service. They serve for the sake of the social need and not, or not simply for their own sakes” (Walzer 1983, 89). In order to achieve full decommodification, on Walzer’s view, the state must act as both purchaser and provider, because anything less than this will result in the transaction being mediated through a money price (with the state either taking bids from rival suppliers, or else providing goods to those willing and able to pay).

But again, it is important to observe how seldom the welfare state actually achieves, or has even aspired to achieve, this sort of decommodification. Even with the NHS, the suggestion that health care is distributed in accordance with need is something of an illusion. It is actually distributed in accordance with a physician’s determination of the patient’s needs. Unlike the U.S., in which patients can directly access specialist services, in the U.K. general physicians act as gatekeepers, limiting access to the system so that more expensive services are available only by referral. (This is a general feature of most ‘needs-based’ systems, such as elder care, in which access is always determined by a bureaucratic assessment of need—see Evans 2016.) But more importantly, complete decommodification is not necessary in order to institutionalize this access principle. The Canadian single-payer health system, in which the state acts only as the purchaser, leaving provision to the private sector, has an access structure almost identical to that of the NHS. Health care is distributed in accordance with need because it is paid for by insurance, and all insurance
(whether public or private) offers indemnification based on the need generated by the insured event (Horne and Heath 2022).

Nevertheless, there is a widespread belief that the welfare state imposes limits on the market, by taking certain goods and both producing and distributing them in accordance with non-market principles. This is often approved of, not just because of a distaste for the commodity form, but also because of its egalitarian distributive consequences. Since the income tax system is progressive, it will typically be the case that the burden of paying for goods that are financed through general taxation will fall more heavily on the rich. This distributive argument typically underdetermines the case, however, because the same objective could be achieved by redistributing income, or by providing individuals with vouchers to access certain goods (Daniels and Trebilcock 2005). And indeed, this is precisely what the state does in many domains (ranging from food security to higher education access). Thus the egalitarian perspective offers no explanation for why the state needs to be more deeply involved on either the purchaser or the provider side of the transaction.

As a result, there has been a certain amount of puzzlement among political philosophers, generated by the fact that money is still changing hands in so many areas of welfare-state intervention, and a great deal of provision is being carried out—and always has been carried out—by the private sector (Radin 1989). These goods (such as public roads) are sometimes referred to as ‘partially decommodified’. But this obviously invites the question, whether the objective was ever to decommodify these goods in the first place, or whether some other normative principle provides a more perspicuous reconstruction of the reasoning that led to them being in the public sector. The practice of partial decommodification is so ubiquitous that it tends to raise doubts about the plausibility of the decommodification standard. It is also worth noting that, unlike the postal service, which enjoys a state-imposed monopoly that prevents private firms from offering competing services, many other goods provided through the TSPM can be ‘topped up’ by private purchase. Typically, the state provides these goods at a very basic level for free, but allows individuals to purchase additional quantities, or higher quality versions, from the private sector. This is unlike goods such as transplant organs, or surrogacy services, in which the state prohibits all commercial transactions (Sandel 2012). By contrast, there is an active market for private health care services and health insurance in the U.K. Similarly, with garbage disposal the state acts as the primary purchaser, but businesses are typically obliged
to contract directly with private firms to secure disposal. The police provide basic reactive security, leaving individuals free to purchase additional security (especially proactive security) from the private sector (Claassen 2011; Schafer and Fehling 2021). Thus these goods are not really being ‘decommodified’ through state involvement.

My own view, which I will mention briefly here, is that the best way to understand all of the primary domains of economic welfare state activity is to see them as responding to one or another form of market failure (Heath 2011). In some cases, such as education, state intervention is motivated by the presence of significant positive externalities (and the tendency of markets to underproduce them). In other cases, such as health insurance, it has to do with information asymmetries (and the fact that the state is not vulnerable to adverse selection). In other cases, it has to do with market power (and the tendency for natural monopolies to form in the private sector). The important point is that market failures can occur on either the demand side, or the supply side, or on both sides of a market. It is only in the latter case that the TPSM is necessary as a response; in other cases state intervention is only required on the purchaser (demand) side or the provider (supply) side, depending on where the market failure is occurring.

IV. Types of Privatization

Given the prevalence of purchaser-provider splits in the welfare state, it seems unlikely that there could be any principled objection to such arrangements (certainly there can be none that appeal to precedent and best practice). Otherwise put, the insistence that all economic activities of the welfare state conform to the TPSM is excessively rigid and rules out a number of highly successful welfare-state arrangements (such as the Canadian health care system, or the U.S. Social Security system). Furthermore, it runs into serious difficulty, and perhaps conceptual incoherence, once the boundary problem is recognized (that is, how far ‘down’ does the TPSM have to go?). Does anyone care that mail delivery vehicles are not being built by government employees? But then does it matter whether they are maintained by government employees?

---

6 Again, I am excluding purely redistributive programs implemented through the tax system and entitlements. The magnitude of the redistribution undertaken is often vastly overestimated by commentators who misclassify state-run insurance plans as redistributive programs.
It is useful therefore to distinguish the TPSM from a purchaser-provider split model (PPSM), in which the state acts either as the primary (or sole) purchaser of a good, or as the primary (or sole) provider of a good. The former can be described as the demand-side public sector model (DPSM), the latter as the supply-side public sector model (SPSM). This then generates a variety of different forms of privatization (labelled 1 through 5 in Figure 1). In principle one could move from the TPSM to a fully private market (FPM), characterized by private provision and funding through direct purchase, although this is relatively uncommon. Most privatizations have involved movement either from the TPSM to a PPSM (that is, the introduction of a purchaser-provider split), or from a PPSM to a FPM (cf. Starr 1989, 24). Many philosophical critics of privatization, it should be noted, such as Alon Harel, Avihay Dorfman, and Chiara Cordelli, appear to be concerned only with shifts from state to private provision, and so raise no principled objections to type 1 and 4 privatization, focusing only on types 2, 3, and 5.\footnote{See Dorfman and Harel (2013) and Cordelli (2020). My central objection to these views is not merely that they ignore certain forms of privatization, but that, in order to rule out the forms of privatization that they do object to, they impose normative conditions on agents of the state so strict that they rule out large segments of the administrative}

---

**Figure 1: Types of Privatization**

---

\[\text{TPSM (state provider, state purchase)}\]  
\[\text{SPSM (state provider, private purchase)}\]  
\[\text{DPSM (private provider, state purchase)}\]  
\[\text{FPM (private provider, private purchase)}\]
five, and so it is worth considering what is at stake in each of these decisions about the role of the state.

Perhaps the closest thing to a type 5 privatization, or a direct movement from the TPSM to a FPM, was the (abidingly unpopular) privatization of residential water services in England and Wales. Historically, residents had been charged a flat rate for water and sewage (that is, regardless of usage), calculated as a fraction of municipal property taxes. The distribution network and supply facilities were owned and operated by the state (as they are still in Scotland and Northern Ireland). Privatization (carried out in 1989 by the Thatcher government) involved not only the creation and sale of shares of these state-owned firms, but also authorization (and encouragement) to institute new billing arrangements. The latter was aimed at promoting the installation of water meters, which allow utilities to charge consumers for the exact amount of water used. This was motivated in part by the fact that the U.K. was lagging most European countries in the transition to metered water supply (although it should be noted that in these other countries there was no privatization of water provision, and so the installation of meters in those countries involved only a type 1 transition from the TPSM to a SPSM).

The more common form of privatization is type 3, a shift from the SPSM to an FPM. Highly publicized cases of this sort involved state-owned enterprises (SOEs), such as airlines, automobile manufacturers, or oil companies, which already functioned very much like private firms in their relations with customers, being sold off to the private sector. Somewhat less common is type 4 privatization, or a transition from a tax-financed to a user-pay model, where the supply is already in the private sector. The most familiar cases of this involve situations in which the provision of some good is subsidized from general tax revenues, which the state then withdraws. The elimination of gasoline and other fuel subsidies are perhaps the best-known instances; the ‘delisting’ of certain services from public health insurance plans would be another example.

state as well, or at least a number of important and well-established practices. Cordelli’s (2020) commitment to non-delegation (121), for example, seems to me to rule out corporatization of state-owned enterprises (SOEs), or the creation of government agencies controlled by a board of directors that is insulated from political influence. Similarly, her concern that representative agents may not act on reasons “external to the state’s public purpose of securing justice” (181) seems to me to rule out public-sector unions, which clearly adopt an adversarial stance toward employment relations, and so cannot be construed as sharing the public-interest orientation of the state. Thus the public state, as she describes it, is not one that exists, even in approximate form, anywhere in the world.
What gives rise to a great deal of confusion, however, is the fact that the term privatization is used not only to refer to these transitions to a fully private market, but also to describe many transitions from a TPSM to either a SPSM or DPSM. These forms of privatization do not result in the formation of a fully private market, because the state still retains an important role on one side of the transaction. Thus the shift away from taxpayer funding toward tuition-based payment is described as the privatization of higher education (type 1), even when ownership remains public. Or the shift away from having city employees collect residential garbage, to the contracting out of this service to a waste management firm, is described as privatization (type 2), even when it remains financed by taxation and free at the point of service. This is certainly the area in which one is likely to find the most anodyne instances of privatization, simply because the state retains the ability to exercise considerable control over the market by virtue of dominating one side of it. At very least, there can be no categorical rejection of such arrangements, simply because the purchaser-provider splits that are created through privatization are already ubiquitous in the public sector.

The major question with type 1 and 2 privatizations is, given that the state is still going to be maintaining an important presence in the sector, why not just stick to the TPSM? In particular, the traditional model is often favored for reasons of distributive justice. On the supply side, government employees are often better paid and enjoy better working conditions, including more extensive unionization, than private-sector employees. On the demand side, funding through general taxation is progressive with respect to income, whereas user payments are not. For this reason, opponents often allege that privatization is favored by the enemies of equality precisely because the TPSM has these progressive features (see, for example, ITPI 2016). More careful analysis, however, reveals a somewhat complicated situation. Some of these distributive justice intuitions opposed to privatization turn out to be invalid upon closer examination, while others, which turn out to be valid, can still be achieved under a PPSM, using different policy instruments. Thus the distributive justice

---

8 As Paul Starr (1987) puts it, these changes “privatize the means of policy implementa-

tion but not the functional sphere of government action” (125).

9 The benefits of working conditions in the public sector are sometimes overstated. For example, low-level employees in the public sector, particularly those in public-service roles, typically enjoy a wage premium over their private-sector counterparts. But this premium disappears as one ascends the organizational hierarchy, so that managers are typically earning less. See Richards (1998).
issue often just muddies the water, making it more difficult to focus on the issues that are actually driving the decision whether or not to privatize.

In the remainder of this section, I will describe some of the factors that have played a central role in these deliberations. I will preface these remarks by setting aside some of the ‘free lunch’ arguments that are sometime made, both by critics and defenders of the welfare state, which suggest that one or the other ownership structure is inherently superior to the other. For example, it is commonly claimed by proponents of privatization that government is necessarily less efficient or less innovative than private enterprise.\(^{10}\) This may be a valid generalization, but it is not a universal law (Estrin and Pérotin 1991). It is certainly the case that highly competitive markets tend to promote firms with efficient internal organizational structures, just as markets with low barriers to entry are very effective at rewarding innovation. But it is not the private ownership structure of firms that is generating these beneficial effects, it is the external environment in which they are operating. Some of the greatest disappointments in the history of privatization have occurred when states sold off assets to the private sector, but failed to create a competitive market for those firms to operate it, leading to the grim discovery that private monopolies can be just as inefficient as public monopolies, and often more infuriating to deal with. Thus the promise of efficiencies arising \textit{merely} from the change in ownership structure winds up being a free lunch (that is, an illusory benefit).

Opponents of privatization often make an opposite, but equally invalid claim, by suggesting that the public sector is always more efficient because it is not burdened by the expense of paying out ‘profits’ to investors. This view remains quite persistent, despite the fact that it is an obvious free lunch. Since every welfare state in the world today runs a substantial budget deficit, every marginal increase in expenditure is financed by debt. And so either way, providers of capital get paid, whether in the form of interest on government bonds to finance public-sector investment, or in the form of profits on shares to finance private-sector investment. Explicit borrowing costs are somewhat lower in the public sector, since the state need not pay the ‘equity premium’ that is enjoyed by shareholders in private firms. But that is because the state is bearing all of the risk in the case of public financing, which means that lenders are being compensated in the form of a guaranteed return on their investment.

\(^{10}\) See discussion in Mazzucato (2015).
Either way, the providers of capital get paid an equivalent return. The alternative, of financing expenditure through increased taxation, is obviously not free either, although the cost is much more difficult to calculate (involving both transaction costs and deadweight losses). Nevertheless, it is difficult to find any estimate of the cost of raising an additional dollar of tax revenue that finds it to be lower than the rate of return on equity investments.

Thus there are no global arguments for or against privatization, each case must be evaluated on its merits. When thinking about these merits, it is helpful to distinguish the type 1 from type 2 cases.

**IV.I. Motives for Type 1 Privatization**

The easiest way to understand the motivation for type 1 privatization, which involves shifting from tax-financing to beneficiary-financing of state-provided goods, is to consider the consequences of the fact that most goods provided by the welfare state are not actually public goods in the Samuelson sense. Because of this, they are not non-rival in consumption, but rather are subject to congestion, whereby one person's use either reduces the quantity available for others or erodes the quality. Obviously burning natural gas, drinking water, using electricity, or living in an apartment makes these goods unavailable to others. Less self-evidently, driving on a road creates congestion that diminishes the value of the road to others; attending an educational institution increases class size, which may reduce the quality of the learning experience; showing up at the emergency room creates delay that may reduce the quality of care for others; and so on. There may be an excellent case for having these goods provided by the state, but when they are free at the point of service it has the potential to generate collective action problems. Because citizens have no external incentive to limit their consumption, they will tend to ignore the negative effects their decisions have on others, resulting in either shortages (and queues) or quality degradation. This can have negative downstream consequences as well. One of the major factors contributing to urban sprawl, for instance, is that not only are most roads unpriced at the point of use, but municipalities often use general tax revenue to pay for expansion of water, sewage, electrical, and transit systems, as a result of which homeowners do not pay anything close to the true cost of low-density exurban development. Under the TPSM, the government, which is supposed to be in the business of solving collective action problems, may find itself creating them instead.
An important motive for type 1 privatization is therefore to resolve collective action problems that have been created by public financing of a rival, or partially rival good. The installation of water meters is perhaps the least controversial instance of this, in part because of the vast amounts of water that are wasted when it is unmetered (which apart from being expensive, can have very negative environmental consequences) (Ornaghi and Tonin 2021). People growing lawns in arid or semi-desert conditions, or hosing down the sidewalk outside their front door instead of sweeping it, are both easily observable consequences of financing water provision through a flat tax on households. Restaurants in California have on occasion been forced to stop offering water to their customers, while farmers nearby waste literally billions of gallons growing crops like rice and almonds in the desert, all because of the government’s failure to charge users for the amount of water that they consume (and a prohibition on reselling it) (Fox 2015).

Along similar lines, there is a fairly compelling argument to be made for road pricing as an antidote to congestion. Drivers in many jurisdictions do actually finance road construction and maintenance through fuel taxes. The problem is that they pay nothing for roads at the point of use, which leads them to ignore the effects that their decisions about where and when to drive have on other users. Back when tolls had to be collected by hand the imposition of a pricing system would slow down traffic, but with electronic billing this obstacle has been removed. And yet unlike most public transit users, who are forced to pay per trip, the payments made by drivers are still funneled through the tax system, in a way that creates a fairly serious collective action problem. Making them pay directly for road use is a way of reorganizing these incentives. (In opposition to this suggestion, one is likely to encounter the egalitarian objection, that usage charges are equivalent to ‘regressive’ taxes, because they cost the poor a larger fraction of their income. This is a misleading characterization, however, insofar as a price is not a tax, and every price in the economy is, in a sense, regressive, since it costs the poor a larger fraction of their income to buy a television, or a slice of pizza, than it costs the rich. The solution, however, is not to give everyone free roads, televisions, or pizza, and simply tolerate the collective action problems that result, but rather to give income assistance to the poor, and charge everyone the true social cost of their consumption.)

The second major argument for type 1 privatization is that the use of general tax revenue to fund provision of a good may result in unwanted
forms of cross-subsidization between different classes of citizens, including some that are highly perverse. For example, it will generally be the case that the lack of user fees, or the presence of subsidies, in utilities and infrastructure will leave those who seek to minimize their environmental footprint subsidizing the lifestyle of those who do not. There is also the concern that if consumption of the good is skewed toward the wealthy, the net effect of public provision on the TPSM may be regressive, even though the tax system as a whole is progressive. The best example of this is university education, which despite generating positive externalities, also produces non-trivial private benefits. The introduction of variable fees for students was intended to eliminate unjustifiable forms of cross-subsidization between different categories of student, as well as between taxpayers and students who derived significant private benefits from their education (Barr and Crawford 2005).

Finally, it is important to recognize that fiscal discipline may be very difficult to achieve under the TPSM because of the ‘soft budget constraint’. Under the very old-fashioned model, public-sector economic activities were simply part of the operations of government departments, often with no clear demarcation between the budget of different components. Thus the Department of Transportation would spend money on highways, rail transit, airports, and seaports, and earn revenue from various sources, such as a gas tax, carriage fees, ticket sales, and docking charges, all of which would simply be absorbed into the department’s budget, then perhaps supplemented with a subsidy from general tax revenue. This made it very difficult to get a clear picture of the fiscal situation with respect to any particular operation, such as a rail service. In order to avoid the rather serious accountability problems caused by this arrangement, the moderately old-fashioned model saw particular operations being consolidated into distinct ‘crown corporations’ or SOEs, which would be assigned a budget and then held accountable for their own expenses and revenues. While in principle this promised to improve fiscal responsibility, in practice it achieved much less than initially hoped. The central problem was that the state had difficulty imposing credible budgetary discipline on these SOEs. If revenue fell short of projections, or cost overruns led to additional expenses on a capital project, the central state had little choice but to cover the shortfall. So while SOE managers were

---

11 Having graduated before the introduction of variable fees, I paid less than $600 per year in tuition to obtain a degree in philosophy from McGill University, while my wife paid the same tuition rate to obtain a medical degree from the same university. It is not difficult to find multiple forms of unjustifiable cross-subsidization in this arrangement.
instructed not to run deficits, there was not much that could be done to deter them from doing so (other than threatening to fire senior managers, which was also for various reasons seldom a credible threat). This created the problem that János Kornai referred to as the soft budget constraint (Kornai 1986).

Because states have improved their capacity to manage these budgetary difficulties, it is easy to lose sight of how serious a problem they became during the early period of welfare state growth. During the 1960s and 70s, SOEs lost extraordinary amounts of money, enough to undermine the stability of several European currencies, as well as threatening the fiscal integrity of the state. In 1975, for instance, the total budgetary deficits of SOEs in the U.K. was equal to 35% of their contribution to GDP (Vickers and Yarrow 1991, 122). Books such as James O'Connor's, The Fiscal Crisis of the State, published in 1973, reflected the widespread perception that the welfare state was financially unsustainable. A great deal of this red ink was due to simple mismanagement. For example, public sector operations were often overcapitalized in comparison to the private sector: airlines had too many airplanes, flying too many empty seats; railroads had too much rolling stock, and kept open too many branch lines, and so on. But as long as SOEs had the capacity to tap into general tax revenue to cover their deficits, it proved extraordinarily difficult to get them to stop losing money. More generally, the public sector has considerable difficulty penalizing failure (for example, closing things down), in part because those who are adversely affected have political avenues of redress.

The primary mechanism through which these problems were addressed initially was ‘corporatization’, whereby SOEs were put in increasingly arms-length relations to traditional line departments, and given their own budgets, allowed to enter legal contracts, assume commercial liabilities, and so on. Many, of course, continued to be funded by the state, but increasingly their budgets became subject to explicit contracting (for example, through ‘pay for performance’ arrangements, in which budgets would be increased only in return for the achievement of specific objectives). The major reason that these arrangements were introduced within the public sector was to enhance financial accountability, and to supplement traditional lines of ministerial authority with more binding directives. Once this was accomplished, however, budgetary discipline could be further enhanced by requiring these SOEs to finance their own operations by charging for their services, an arrangement that in many cases
involved type 1 privatization. Such an arrangement can be used, for example, to force SOEs to finance their own capital expenditures from the revenue stream they can be expected to generate (something that officials acting in an oversight capacity, looking at an expenditure request, are often at pains to assess). Thus type 1 privatization wound up being used often as a way of imposing, if not a hard budget constraint, then at least a firmer constraint on SOEs.

It is important to emphasize that there are better and worse ways of doing all of this, and there are many examples of privatizations motivated by one or another of these concerns that worked out very poorly. What I have described are simply three problems associated with the financing of state activities through general taxation: it may generate collective action problems (which manifest themselves in the form of wasteful consumption, congestion, quality degradation, and queuing); it may produce unwanted cross-subsidization among different categories of citizens; and it may give rise to a soft budget constraint that encourages wasteful public expenditure. In certain cases, it may be possible to attenuate or resolve these problems by imposing payment at the point of consumption, or otherwise transferring the cost of providing the good to its beneficiaries. As a general principle, one can say that with state programs that are primarily motivated by the goal of correcting market failure, if it is not overly costly to identify the individuals who are benefiting from the program, if the magnitude of that benefit can be estimated with some precision, and if there are no overriding reasons of fairness not to do so, then the state should charge those individuals for the benefit that they receive.\[12\] These are the cases where type 1 privatization is justifiable.

IV.I. Motives for Type 2 Privatization

The second, somewhat more controversial form of privatization involves cases in which the state continues to fund the provision of some good through tax revenue, but instead of having the actual provision done by state employees, it contracts with either private individuals or organizations to do the provision. Type 2 privatization therefore involves a shift in the public-private boundary, so that goods that had once been made internally are now bought from external sources. The most important thing to recognize, when it comes to the ‘make or buy’ choice, is that the

\[12\] An example of a fairness concern that might arise would be if one could identify some of the beneficiaries, but that others who received the same benefit could not be identified, and so the state wound up only charging the former group.
state is often faced with very unattractive options on both sides. The problems with state employees will be discussed below; for now I would like to draw attention to the fact that governments are in many ways handicapped compared to private corporations when it comes to contracting out, because of the extremely onerous constraints that are imposed on procurement. Keeping in mind that practically nothing is public all the way down, every government department and SOE purchases a large number of inputs from the private sector, in the same way that private corporations maintain a large number of supply relationships. The state, however, has enormous difficulty getting good value for its money from these relationships, in part because it is prevented by procurement rules to developing strong relationships with suppliers (Kelman 2002).

Although technically the relationship between a firm and its suppliers is a market exchange, in practice private corporations often cultivate highly cooperative relationships with their long-term suppliers. A supplier may know, for example, that if it overperforms with respect to a contract (for example, delivering goods early or at lower cost, suggesting a more technologically efficient solution, etc.) it will be rewarded through preferential access to future contracts. Government, however, is usually constrained by the requirement that it take the lowest bid on any tendered contract. As a result, the state may find itself obliged to accept a bid from a firm that it considers exceedingly unlikely to fulfill the contract, over another firm that has performed exceptionally well in the past. For the same reason, firms have no incentive to overperform, because they know that it will gain them no consideration in the future. As a result, they tend to behave very legalistically when it comes to performance, providing no more than the letter of the contract obliges them to. The famed capacity of the private sector for innovation and creative thinking is often applied to the task of cynically exploiting government contracts.

Because of this, the state often finds itself in a situation in which it would like to contract something out, but is unable to do so because it cannot secure good value. If the contract is insufficiently complete, contractors will exploit ambiguities, but making it complete would impose excessive rigidity. For example, the public sector tends not to have very sophisticated information technology competence in-house, in part because it cannot compete with the private sector for talent (Kelman 1990). As a result, when it wants to computerize some process for example, switch to electronic record-keeping), it must put out a contract. The problem is that procurement regulations often do not permit the state to
specify its needs in outcome-based terms (‘we would like to solve problem x’), it must instead specify in considerable detail how it would like the problem solved, and precisely how much it is willing to pay. Unfortunately, it often lacks the technical expertise to know how the problem should best be solved. Thus it may find itself unable to tender a contract, because it cannot articulate its needs at a level of specificity that will both satisfy the rules governing the procurement process and protect itself from exploitation by contractors. Ironically, this creates an incentive to privatize government agencies that purchase high-value inputs, precisely because this will free them from government procurement regulations, allowing them to get better value out of their supply relationships.

Attentive conservatives have in fact noticed that procurement regulations are a major barrier to their project of achieving meaningful reductions in the size of the state. Some have argued that instead of engaging in top-down privatization of traditional public services—like selling off the local water treatment plant—which tends to evoke political resistance, conservative governments should instead focus on improving the efficiency of procurement practices, which would in turn promote a bottom-up privatization of state operations, as public managers increasingly find themselves able to secure better value by contracting out more services.

The problem with this strategy is that it is practically impossible to improve procurement practices without increasing the discretion of state officials, which many conservatives are also opposed to, in part because they are suspicious of the motives of public officials. As a result (and contrary to the narrative of rampant neoliberalism), conservative parties have not achieved much success at solving the basic conundrum of government procurement, and have therefore done little to incentivize bottom-up privatization.

When one turns to the other option, of doing in-house provision, the choices may also be unattractive, but for quite different reasons. Although partisans of the welfare state have been extremely reluctant to admit it, public sector unions are the source of an enormous number of difficulties that afflict in-house provision. Unions were developed as a

---

13 One of the most striking omissions in Cordelli (2020) is that she makes no mention of public sector unions. She assumes that state employees share the public commitment to achieving justice, neglecting the fact that they often obey orders (such as ‘work to rule’ instructions) from organizations that stand in an explicitly adversarial orientation toward their employer, the state. It is difficult to see how the profit orientation of a private firm could be intolerable, and yet the activities of unions, which are committed to maximizing benefits to their members, could be permissible.
response to the adversarial labor relations of market contracting in private corporations, and so the importation of those adversarial practices into the public sector has been a poor fit in a number of ways. Superficially, the major problem is that, unlike in the private sector, in which unions are forced to moderate their demands in order to ensure the ongoing viability of their employer, public sector unions need not worry about the state going bankrupt. As a result, they are able to engage in much more aggressive rent-seeking, which results in many countries experiencing more labor militancy, and a great deal more work disruption, in the public sector than in the private. The only real constraint on this is the threat of back-to-work legislation, or revocation of the right to strike (for instance, through designation of a class of workers as ‘essential’). This has a number of peculiar effects, including the fact that it may make unions more obstreperous in dealing with left-wing governments, whom they expect to be more reluctant to enact such legislation. The most unfortunate effects, however, arise at the municipal level, since these governments seldom have the power to enact back-to-work legislation, and so must appeal to a higher order of government to resolve recalcitrant labor problems. In countries such as the U.S., where municipal politics is structured by the same party divide as state and federal politics, this assistance may not be forthcoming for a variety of reasons, include baser forms of partisan politics. Thus municipal governments often have the greatest incentive to contract out various forms of labor, because they are in such a weak position when dealing with public-sector unions.

Sentimental attachment to the labor movement has made both left-wing activists and academics extraordinarily averse to criticizing public-sector unions, or to pointing out the various ways that their actions have contributed to the decline of the socialist project. Despite widespread mobilization against police violence in the U.S., for instance, it took progressives a very long time to acknowledge that police unions were an enormous part of the problem (Levin 2020). Similarly, considerable apprehension has been expressed about the role that corporations running private prisons might play in pressuring states to expand incarceration, with total disregard for the role that (far more powerful) prison guard unions have played in lobbying for the exact same goals (Pfaff 2017). Teachers’ unions in the U.S. also served as a major obstacle to school re-opening in the wake of the COVID-19 pandemic, leading to large-scale flight to private schools (which remained open) in many regions (Marianno et al. 2022). And while there has been an extraordinary amount of attention paid to
the influence of corporate interests in the political process, the problem of capture of social-democratic parties by public-sector unions has been relatively neglected (Barber 2015).

With respect to privatization, perhaps the most dramatic example of public sector unions undermining the case for state ownership is the degradation that has occurred in public housing projects in North America. The issue here is not just wages but also the extremely rigid work rules that unions favor. Keeping older buildings in a state of good repair requires a complex combination of skills. Private landlords will typically employ a ‘handyman’, who is able to carry out most small maintenance and construction tasks, including plumbing, electrical work, carpentry, and painting. Unions, by contrast, will usually insist on a strict separation of trades. The result, as one official at the Boston Housing Authority described it, is that even small repair jobs may require appointments with several different trades (“if you had a leak under the sink, a laborer had to go and break open the wall. A plumber would have to get in there and fix the pipe. A carpenter had to come back up and frame the hole back in. The plasterer would have to plaster. And the painter would have to paint”—Moore 1995, 199). These onerous work rules are a major reason that public housing systems across North America have a multi-billion dollar backlog of repairs. Budgetary authorities, however, have no confidence that sinking a few more billions of dollars into these operations will actually result in many repairs getting done. As a result, the housing stock is simply allowed to degrade until eventually it becomes uninhabitable. Under such circumstances, there is a strong argument to be made for privatization (combined, as appropriate, with a voucher scheme allowing low-income residents access to rental housing on comparable terms).

If the practical consequence of public provision of housing is the degradation and eventual destruction of entire neighborhoods, it is unreasonable not to explore other ways of securing access to those goods.

---

14 This account reminds me of the time that I requested a computer for my office, which at my former university needed to be acquired through the internal purchasing system. After several weeks the machine finally arrived, with strict instructions that I was not to open the box or do anything to set it up. Union regulations required that installation only be performed by an appropriately certified technician. I therefore had to make an appointment and wait another full week before a technician could come to my office, remove the machine from its packaging, plug it in, and perform the generic Windows set-up. It is easy to be distracted by the comedic aspect of these kabuki-like work rules, and to lose sight of the fact that, when practices like this become generalized they constitute a serious handicap for the public sector.
The example of public housing is somewhat anomalous, in that there were not really significant market failures in this sector. While private landlords exercise some market power over tenants, because of the costs involved in moving, the primary motivation for the development of public housing was always to improve affordability, which is an essentially redistributive objective. The challenge with redistributive programs is then to justify in-kind provision, since if the market is reasonably efficient, the egalitarian objectives can be achieved through monetary transfers. Thus certain type 2 privatizations are motivated by the realization that, in order to ensure that all citizens have access to certain goods, it is not necessary for state employees to be providing them directly. The Supplemental Nutrition Assistance Program in the U.S., for example, is run by the Department of Agriculture because in the distant past the U.S. government used to take surplus agricultural product and transfer it directly to the poor. Food stamps were introduced in 1939 in order to give recipients greater freedom of choice, while still focusing on the disposal of agricultural surplus (as part of a broader price-management scheme). This evolved into the contemporary program, which provides recipients with an electronic card that can be used to purchase most food items, and in some cases inexpensive restaurant meals. Most countries take the easier route of simply providing the poor with cash payments, with the expectation that they will use a certain fraction of it to purchase food. In each case, Western states have essentially gotten out of the business of providing food directly to the poor; food security is achieved by providing a means-tested transfer, leaving the provision of food to the private sector.

Housing projects in the U.S. are also peculiar in that they are administered municipally, but derive significant funding from the federal government (from the U.S. Department of Housing and Urban Development). This illustrates another important point, which is that theorists often refer to ‘the state’ as though it were a unified organization, ignoring the fact that authority is typically fragmented between different levels of government. This is particularly important in federal states, where regional governments often wield enormous power, and yet act in an entirely strategic manner toward other regions in the confederation, as well as the federal government. Thus the idea that government officials act in ‘the public interest’ harbors a significant ambiguity, since the public interest at lower levels of government is typically identified with a regional interest. Many state or provincial politicians do not even attempt to sustain the pretense
that they care about the interests of the country as a whole, because of the electoral advantages that can be obtained from adopting an antagonistic stance toward outsiders.

This can become a serious problem when SOEs belong to these lower orders of government. For example, elected representatives from coal-producing regions of the U.S. can be counted on to oppose any effort at climate change mitigation. This is typically taken to be a consequence of successful lobbying and political donations by the coal industry. But now imagine if state governments owned the coal companies! This would, if anything, make the problem worse, since public managers have internal access channels (and are immune to lobbying restrictions imposed on private firms) that can make them an incredibly powerful force within government. One can see this in Canada, where provincial governments have jurisdiction over natural resources, which gives them both control over resource royalties and the freedom to develop SOEs in the mining sector. This has made provincial governments extremely aggressive in defending the interests of the mining industry, with a number of consequences, including considerable success in opposing federal environmental regulation. The influence of the Quebec government, for example, in defending the interests of the Société nationale de l’amiante is one of the reasons that asbestos mining continued in Canada for so much longer than in Europe and the U.S. (Shields 2012).

All of these are examples of the way that the accumulation of significant resources and economic power within the state can be, in effect, corrupting, benefiting constituencies whose interests are poorly aligned with the general public interest. This is a reflection of the fact that public management—which has as its central priority maintaining that alignment—is extraordinarily difficult, being subject to a great many more constraints than private management. As a result, cases may arise in which it is easier to maintain that alignment through external contracting than through internal supervisory arrangements. Under such circumstances, there may be a compelling case to be made for type 2 privatization.

V. Degrees of Privatization

In addition to the five types of privatization described in the previous section, it is important to observe that privatization need not be an all-or-nothing affair. It is possible to partially privatize government operations on the supply and the demand sides of a transaction, or both. Consider, for example, type 5 privatization, which involves a shift from the
TPSM to a fully private market. As already noted, states only rarely enforce a monopoly in the provision of a good, and so private markets typically co-exist alongside state provision. In the standard run of cases, the state provides the good at a relatively low level, but makes it universally available. Individuals who want more can then purchase additional quantities through private contracting. In such markets, type 5 privatization sometimes occurs by attrition, as the state simply reduces its funding for provision of the good, so that the private sector takes on an increasing share of the burden. For example, there was clearly a partial privatization of policing (or security) in the U.S. with the formation of gated communities, since the security guards hired by homeowner associations were not just supplementing the police, but rather were replacing them in the performance of certain tasks (Jones and Newburn 2006; McKenzie 1994; Stark 2009, 29–42).\footnote{Note that gated communities suffer some unfair criticism on this score, since they are not doing anything different from what residents of high-rise apartment buildings have long done when they post a private security guard at the downstairs entrance. The only difference lies in the spatial configuration of the residential units benefiting from the private security.}

Similarly, there are many cases in which the state subsidizes the provision of some good, but where consumers must still pay a fraction at the point of delivery. Many public services have user fees (for example, public swimming pools and sports facilities, summer camps and after-school programs, university tuition), which do not cover the full cost of providing the relevant good. It is possible then for the state to shift the public-private balance by drawing down the subsidy, so that a larger fraction must be financed by these fees. This can be seen as partial privatization of the type 1 or 4 variety.

There are many cases in which this sort of privatization is unjustifiable or even pernicious. In cases where the state is the sole provider of a good, underfunding will typically generate complaints about quality degradation or increased queuing. The availability of a private alternative, by contrast, typically insulates wealthier individuals from the effects of declining quality or availability in the public sector (they are more likely, in other words, to exercise an exit option rather than voice). This leaves only relatively powerless citizens consuming the public version of the good, people whose complaints can more easily be ignored. Thus partial privatization may constitute a form of ‘privatization by stealth’, which winds up being the course of least resistance, politically, only because the costs are borne disproportionately by the disadvantaged.
At the same time, it is important to recognize that there are other cases in which it is perfectly appropriate for the state to adjust the public-private ratio in the provision or subsidization of certain goods. On the subsidization side, circumstances that warrant it include those in which severe congestion or queuing has developed in the public sector, or when users have figured out a way of appropriating for their private benefit a larger fraction of the value produced by the good. For example, with the increased demand for ‘quants’ in investment banking, university mathematics programs found themselves teaching more and more students who were preparing for a career in finance, rather than science. Introducing a separate finance stream, and charging students higher tuition for those degrees, struck many as a reasonable response. On the provision side, partial privatization is often used to introduce competition, which puts pressure on public-sector organizations to improve their performance. Michael Barber, for instance, estimates that contracting out 10–15% of operations in some domain is often enough to generate noticeable improvement in the segment that remains in the public sector (Barber 2015). This is partly because the availability of a private-sector alternative generates benchmarks, which allow for the assessment public-sector performance, and partly for its incentive effects, as public-sector organizations improve their performance (and scale back labor militancy) in order to avoid being privatized.

What is less often realized is that when the state sells off certain assets to the private sector, it may not do so entirely, but will retain some fraction of the ownership, often enough to continue exercising control over the enterprise. In many cases the state will also create for itself a special class of shares (that is, a ‘golden share’ arrangement) that allow it to exercise control regardless of what happens with the shares held by private investors (Bortolotti and Faccio 2009). For example, the sale of Canada’s national oil company, Petro-Canada, in 1990 was by far the most controversial privatization undertaken by the government of Prime Minister Brian Mulroney. And yet the Canadian state actually retained a 19% ownership stake in the firm, and placed conditions on the traded shares that prohibited any individual or organization from acquiring more than 10% ownership. (Similarly, despite being routinely classified as an SOE, British Petroleum was never entirely owned by the British Government. When the first round of ‘privatization’ occurred in 1979, the Thatcher government merely sold 5% of the company’s shares, drawing down state ownership from 51% to 46%.) States also often direct public pension funds
or banks (such as the public investment bank Bpifrance) to invest in the newly privatized corporations, so that private ownership is actually something of an illusion.

This type of mixed ownership is possible because of the prior process of corporatization, which transformed the relationship between many SOEs and traditional line departments. These reforms were undertaken with the primary objective of improving accountability within government. One of its effects, however, was to radically reduce the difference in governance structure between public and private enterprises, which in turn made mixed ownership feasible. Most obviously, corporatization disembedded economic operations from traditional departments, where they were often under direct ministerial authority, and placed them (typically) in a special-purpose statutory corporation, under the authority of a more-or-less independent Board of Directors. And instead of being given a complex mandate, with multiple benchmarks of success, managers were instructed to operate under standard business principles, including in some cases the goal of maximizing profit (especially when demand was fully private). The pursuit of ‘social objectives’ was secured, not by assigning additional objectives to managers, but rather by contracting with the SOE, in effect paying it additional sums to pursue goals such as regional development, hiring disadvantage workers, or keeping the cost of certain goods artificially low.\(^{16}\)

The latter point merits brief comment, since critics often observe that after privatization states are still forced to engage in complex contracting and regulatory intervention in order to control the private enterprises that are supplying the goods that were formerly provided by state employees. This is particularly noticeable in cases in which states privatize operations but fail to create a competitive market, and so are forced to regulate, and in some cases supervise, the newly-created private monopoly. This has given rise to some notoriously opaque and ill-conceived contracts. Critics were certainly correct in observing that the more enthusiastic proponents of New Public Management reforms vastly overestimated the capacity of the state to formulate complete contracts with private suppliers, mainly because they ignored the information asymmetries involved in the provision of many traditional state services (Heath and Norman 2004).\(^{17}\) All of this goes to show how easy it is to underestimate

\(^{16}\) For discussion, see Heath and Norman (2004).

\(^{17}\) Simplifying somewhat, the major issue is that public organizations are often charged with pursuing multiple goals that are not entirely consistent with one another. This in
the agency problems involved in contracting with private firms. At the same time, it is also easy to underestimate the agency costs involved in traditional state employer-employee relationships. Critics of privatization sometimes write as though, under the TPSM, elected officials simply gave orders, which were then faithfully carried out by a cadre of loyal civil servants (Farrell 2019). In fact, these internal channels of authority are often ineffective or non-existent, as a result of which the state already engages in a significant amount of regulation and contracting in dealing with its own branches and operations (Hood et al., 2003). For example, the Minister of the Environment cannot simply order public utilities to reduce their pollution output. Emissions reduction from public utilities is achieved in the same way that it is with private utilities, through regulation. Thus the difference in ownership between SOEs and private corporations does not make as much difference as one might think, when it comes to asserting control over these organizations.

As a result, when the state sells off some fraction of an SOE, this does not necessarily change the level of control that it can exercise over the firm, especially when the state retains a controlling stake and so is in a position to make all appointments to the Board. Partial privatization, in this case, is often just a way of financing capital investment by drawing in equity instead of issuing debt. (Similarly, many public-private partnership arrangements are designed to keep large capital investments, especially in infrastructure projects, off the state’s balance sheet.) In some of these cases privatization more closely resembles an accounting shenanigan than an actual sale. Consider the situation of the British Water Authorities prior to privatization. The British state had, over the course of decades, failed to make appropriate investments in upgrading its treatment facilities and distribution network, as a result of which it was facing the need to make massive capital investments in its water services. Apart from political ideology, it was this need for capital that served as a major motive for privatization. The standard way for the state to finance this sort of investment would have been by selling bonds. The problem for the government of the day was that issuing bonds shows up as government debt, and the Thatcher government could not tolerate the optics of increasing the annual deficit. Equity, by contrast, can be booked as revenue, not debt, and so by attracting equity investment the government was able
d
\[turn \text{ requires making trade-offs, where finding the optimal balance requires judgment, the outcome of which cannot be specified contractually.} \]
\[18 \text{ As a result, it would seem that the principled objections to privatization advanced by Harel, Dorfman, and Cordelli would not apply.} \]
to create the impression of having reduced its deficit. Of course, in the case of water the British state sold off the assets entirely, but had it merely sold off a fraction, sufficient to cover the capital cost of the new investment required, it would have managed to borrow millions of pounds, while at the same time reducing the state’s official debt, and without actually relinquishing any effective control over the water service. (A cynical observer might suggest that the government of France has on occasion done precisely this, using partial privatization as a way of borrowing money from the pension fund of public-sector employees without having to book the loan as debt.)

VI. TECHNOLOGICAL CHANGE

The most highly valued economic activities of the welfare state occur in areas in which the private alternative is compromised by some form of market failure. The concept of market failure, however, is not structural, but is always relative to the current state of technology. What typically causes the market to fail is not the impossibility of solving a particular organizational problem privately, but rather the cost of doing so. For example, it is commonly claimed that negative externalities arise because of incompleteness in the system of property rights. It is, however, seldom impossible to establish property rights in areas in which they currently do not exist, such as over the open seas or the atmosphere, it is just prohibitively expensive to do so. Similarly, insurance markets are often undermined by asymmetric information, but there is usually some price at which the desired information could be obtained. The market fails simply because the cost of organizing the transaction exceeds its economic value, and so private parties do not bother undertaking it. Or in the case of partial market failure, the cost results in inefficiently low levels of production of the good.

Just as technological development can change the profile of transaction costs faced by firms, creating an incentive for them to readjust their boundaries, technological development can resolve old and create new market failures. Most obviously, it can allow competitive markets to develop in sectors that had previously been natural monopolies. Telephone service, for example, had once been a natural monopoly, because it was necessary to connect each building up to a physical network of transmission lines. With the development of wireless service, the cost of developing a competitive market (that is, redundant geographical coverage) declined significantly, resulting in most states opting for a regulated
competitive market over monopoly state provision. Even in France, where the state went to considerable effort to ensure that France Télécom, reborn and rebranded as Orange, would retain a dominant market position, no attempt was made to prevent the emergence of a competitive private market. (Orange was also subsequently semi-privatized, with the French state maintaining a controlling stake.)

More subtly, the development of information technology has radically reducing the cost of private-sector contracting. The decline of the mid-20th century conglomerate firm is largely due to a decline in the cost of market contracting relative to managerial organization. Similar forces have affected the public sector. Although the cost of collecting taxes has declined, the cost of private transactions has declined a great deal more, which has eroded the comparative advantage of the public sector. In particular, keeping track of people’s consumption and then billing them has become dramatically less expensive than it once was. This was the major factor making it possible to split distribution from provision in physical networks such as the electricity grid or natural gas pipelines. As a result, it became possible to create a competitive market for electricity generation, while keeping electricity distribution a publicly owned monopoly. The development of water meters, or time-of-use pricing for electricity, has also become feasible through a dramatic decrease in the cost of collecting data from locally installed devices. And finally, computerization has generated a significant decrease in the transaction costs involved in running auctions, which has in turn enhanced the state’s ability to generate competition among private providers in the absence of conventional markets (such as with electricity supply) (Maurer and Barroso 2011).

In other cases, the general cheapening of commodities has resolved problems involving the underproduction of positive externalities (or non-excludable goods), and thus reduced the need for state involvement. During the first half of the 20th century, for example, the city of Montreal made a major investment in the construction of public baths, aimed primarily at ameliorating the quality of life and social integration of poorer immigrant and working-class communities, who had no hot water in their homes (Labonne 1996). With improvements in indoor plumbing, however, most of these baths fell into disuse (and municipal bylaws no longer permit the construction or rental of living units without hot water and private bathing facilities). As a result, the baths were either sold off or transformed into swimming pools. Although there was some mobilization aimed at saving the buildings—some of which were quite beautiful—no
one opposed the actual privatization, simply because there was no longer any case to be made for public baths. One can imagine something similar happening with public libraries and books, once distribution becomes fully electronic.\footnote{And so, while I love libraries almost as much as Martin O’Neill, I am not persuaded by his claim that they are public because “there is something transcendent, something deeply valuable, indeed almost miraculous, about a shared public space devoted to a role as a repository for human learning and memory” (O’Neill 2020, 187). I can imagine a similar paean having been written 80 years ago about the solidarity-enhancing qualities of communal bathing, and yet no one would think, in the present day and age, that this constitutes a strong argument for public baths.}

These examples of technological change resolving market failures should not lead one to assume that change always runs in the direction of greater privatization. In some cases, technology has created market failures where previously there had been none. For example, physical limitations often made it possible to bundle together products in a way that facilitated private provision of certain positive externalities. Newspapers are perhaps the best example. A traditional newspaper derived significant revenue from its classified advertising section, which it then used to subsidize other aspects of its operations, most notably the news section. The development of the internet made this sort of bundling infeasible, and paid classifieds were soon replaced by online platforms offering the service without charge. This represented an obvious gain for consumers, except that it removed the subsidy to news production, which has produced an increasingly conspicuous market failure in the production of ‘public interest’ journalism, particularly at the municipal level where many local newspapers have disappeared.

Technology has also allowed the formation of new natural monopolies, particularly through the network externality effects that are highly visible on social media. In many cases, the solution lies in the development of open-source versions (such as Linux in the domain of operating systems, or Apache in web servers) and so state ownership and provision are not required. The point is that conditions are dynamically unstable, and so it is important to maintain ideological flexibility when considering the overall merits of state ownership in particular sectors. There are, as they say, many ways to skin a cat, and so when circumstances change the best way to achieve certain state objectives—whether best articulated in the language of social welfare or of social justice—may also change. In certain cases, privatization is justifiable because state ownership or financing is simply not required any longer to achieve those results.
**VII. CONCLUSION**

My objective in this discussion has not been to defend privatization as a general panacea for the various frustrations that citizens experience when dealing with government. The narrow objective was merely to present a counterpoint to the extremely negative portrayal of privatization that is common in the philosophical literature, which arises primarily from a focus on worst-case scenarios. This is not to deny that there have been many terrible, ill-considered, or even flagrantly corrupt privatizations. My idea, however, was to present a few best-case scenarios, in order to show that privatization of the economic functions of the state is not *intrinsically* objectionable, even though it may be, for various contingent reasons, highly undesirable. Examples of anodyne privatization would include the selling of SOEs in sectors that were once natural monopolies, but where technological change has made it possible to organize appropriately competitive markets; the imposition of user charges for publicly provided goods that are experiencing quality degradation from open access (such as electronic tolls on highways that suffer from severe congestion); and the contracting out of public services in cases where relatively complete and enforceable contracts can be written (such as municipal garbage collection). Thinking about these cases alongside the undesirable ones allows one to form a conceptual framework, or a grid of sorts, that will make it easier to situate and to develop a more nuanced analysis of any proposed privatization.

A great deal of the political rhetoric surrounding privatization is a hold-over from an earlier era, in which social-democratic parties were still committed to the gradual nationalization of the entire economy. It is important to recall, for instance, that during the Thatcher years the British Labour Party still had the controversial ‘clause IV’ in its constitution, which expressed a desire for “common ownership of the means of production, distribution, and exchange” (Gani 2015). If this is the objective, then of course any privatization whatsoever constitutes a setback that must be reversed at the earliest opportunity. But since that time, most progressives have come to accept that total state ownership and control of the economy is neither feasible nor desirable. Indeed, most now accept some version of the provocatively ambiguous German Social Democratic Party slogan: ‘As much market as possible, as much state as necessary’ (Crouch 2021).
But if the state is no longer aiming to take over the entire economy, the most pressing question for the democratic welfare state becomes one of determining which parts should still be owned and/or controlled. My own view, which I have defended at length elsewhere, is that state involvement in the economy should be guided primarily by the norm of efficiency, which is to say, the objective of correcting market failure, and in cases in which it cannot be corrected, the substitution of state purchasing and/or provision for market contracting (Heath 2020). This is a view that is commonly held by economists, and much less commonly held by philosophers and political theorists. The received view among that latter has been that state involvement in the economy is primarily motivated by the goal of promoting greater equality. It follows from this that a certain amount of inefficiency in state provision is tolerable, if the arrangement is still beneficial from the standpoint of distributive justice. (To give a concrete example, proponents of such a view may oppose road tolls, even though the failure to price vehicle travel generates serious collective action problems, because it nevertheless makes roads accessible to both rich and poor on equal terms.) The inefficiency, from this perspective, is a cost worth bearing for achieving the more egalitarian outcome.

It is this last conclusion that I would like to cast doubt upon, or at least problematize, in my concluding remarks, by drawing attention to some hidden costs associated with state involvement in economic activities in which it does not enjoy a comparative advantage over the private sector. It is helpful to begin by identifying areas in which the state is an essential actor, in the sense that if it does not provide a good, no one else is going to step forward and do so. (This is the case, for example, with various elements of the social safety net, where if the state does not provide these forms of security to citizens, no private insurance company is going to start selling policies that offer indemnity against those risks.) There are other areas in which state involvement is obviously optional, in the sense that withdrawal from the sector would create a vacuum that private corporations would rush to fill. In between are areas in which state involvement is merely desirable, in the sense that it provides an improvement over what the private sector would have provided.

The problem with maintaining state involvement in sectors in which its activities are purely optional, or do not have a strong efficiency rationale, is that there are often private-sector alternatives available, and it is in many cases easy for citizens to compare the performance of the state to that of private firms. This is a comparison that seldom works out
favorably for the state. There are complex reasons for this, but one of the most important is that providing polite and responsive customer service is an extraordinarily difficult managerial challenge, which many highly-motivated private firms fail to do well, but which public sector organizations, facing only dull pecuniary incentives, not to mention an entirely unionized workforce that cannot easily be penalized for mere rudeness, unhelpfulness, or obstructionism, almost inevitably do badly. Unfortunately, most citizens formulate their opinions about the qualities of the state as an institution on the basis of these interactions with street-level bureaucrats (Rothstein 2011). They also tend not to distinguish essential from optional state activities, and so to the extent that they develop a negative attitude toward state involvement in the economy, tend to apply it to all its activities. As a result, negative experiences with a state-run airline or hotel can easily translate into opposition to state pensions or health care, becoming in turn easy fodder for right-wing politicians with an ideological hostility to government activity in all domains. As a result, when the state does a bad job in areas in which its involvement is optional, it may erode the legitimacy it requires to carry out tasks in which its involvement is essential (Rothstein 2011, 93).

As the experiences of the 20th century have shown, successful government programs, such as Social Security in the U.S., the National Health Service in the U.K., or Medicare in Canada, which both create significant value and promote greater equality, are positively loved by citizens, and may become an important locus of solidarity and national identity. By contrast, poorly run programs, which frustrate citizens and provide undeserved economic rents to insiders, are detested, and have the capacity to generate blanket hostility toward government. Thus it is important that supporters of the welfare state maintain a flexible and intelligent approach to thinking about the role of the public sector, and encourage it to focus its efforts on areas in which it has a demonstrated capacity to create value. Within these domains, there may be many opportunities to finance or distribute these benefits in a way that will reduce overall inequality. The important point is that these egalitarian ambitions be pursued without sacrificing the positive sum character of the overall program.

---

20 An attitude immortalized in Ronald Reagan’s quip about the nine most terrifying words in the English language: ‘I’m from the government, and I’m here to help’. Pedants have pointed out that this is, in fact, eleven words.
Finally, it would be helpful if supporters of the welfare state, instead of simply opposing privatization politically, were to pay more attention to improving public administration in ways that would make privatization less attractive organizationally (mirroring the way that clever opponents of the welfare state have focused on improving procurement, in order to make contracting out more attractive). Many supporters of the welfare state have been too quick to accept the refrain that problems in the public sector can all be blamed on miserly governments imposing budgetary restrictions. As Michael Lipsky has observed, budgetary shortfalls are a structural feature of most public programs, because demand typically expands to absorb any increase in supply. Thus it is important to be judicious in evaluating claims about underfunding. While it is true that some problems can be solved by spending more money, questions of privatization often turn on the relative efficiency of different organizational forms, which means that spending more money on public operations often just bolsters the case for privatization. The fact that public-sector organizations can simply demand more money from the public purse is one of the major problems with public-sector organizations, that makes them so difficult to manage. The arrangement we should be aiming for is one in which the state chooses carefully the tasks that it is to perform, in such a way as to ensure that what it does, it does well. In order to achieve this, the state must obviously be prepared to stop doing—which is to say, to privatize—that which it does badly, or that others have learned to do better.

REFERENCES

21 “A distinct characteristic of the work setting of street-level bureaucrats is that the demand for services tends to increase to meet the supply. If additional services are made available, demand will increase to consume them. If more resources are made available, pressures for additional services utilizing those resources will be forthcoming” (Lipsky 1980, 33).


**Joseph Heath** is Professor in the Department of Philosophy at the University of Toronto.

Contact e-mail: <joseph.heath@utoronto.ca>
Privatization, Structural Dependence, and the Problem of Legitimacy

CHIARA CORDELLI
University of Chicago

Abstract: How should a normative evaluation of the merits and demerits of privatization proceed? In response to Joseph Heath’s approach to this question, I first argue that the difference between core and economic functions is not as relevant to establish the limits of privatization, as Heath suggests. I claim, second, that Heath problematically neglects the structural and aggregative effects of privatization. An instance of privatization that, if analyzed in isolation, is anodyne may no longer be so when seen as a further contribution to an already expansive process of privatization. Finally, I argue that Heath fails to consider the risks that pervasive privatization poses in terms of democratic legitimacy, by giving rise to a situation where, on the one hand, citizens are dominated by a privatized state and, on the other hand, the state itself is dominated by private actors.

Keywords: privatization, legitimacy, efficiency, domination, structural dependence

JEL Classification: DS, H4, I13

I. INTRODUCTION

How should a normative evaluation of the merits and demerits of privatization proceed? In response to this question, Joseph Heath (2023) has advanced three main theses. First, what we may call an asymmetry thesis: “as a prelude to any detailed discussion of the merits of privatization”, he argues, “it is essential to distinguish core functions of the state from what are often described as ‘economic’ activities” (26). While, as far as the coercive apparatus of the state is concerned, “it would not be surprising to discover that there are strong objections to privatization, stemming from basic philosophical principles” and, indeed, it would be “courting disaster to begin replacing elements of that apparatus with private organizations” (28), by contrast, within the realm of economic activities, there
are cases of anodyne privatizations, understood as “privatizations that no reasonable person could object to” (26).

Second, and relatedly, Heath proposes a case-by-case thesis. Since, he argues, there are no global arguments for or against privatization, “each case must be evaluated on its merits” (41). The (de)merits of privatizing the management of highways may be very different from the (de)merits of privatizing the provision of childcare. Therefore, the only plausible way of assessing the justifiability of certain privatization decisions is to analyze the specific costs and benefits involved in each instance of privatization, assessed in its own terms.

But how should we think about the ‘merits’ in relation to which each case of privatization must be so judged? Heath’s answer—his third thesis—is in line with mainstream economic theory: “state involvement in the economy should be guided primarily by the norm of efficiency, which is to say, the objective of correcting market failure” (60). When the market fails on both the demand and the supply side, then the state must act as both a purchaser and as a provider. But in most cases market failure occurs on one side only, and it is on that side only that state intervention is required. This may sound slightly confusing since, as Heath himself later shows, many types of privatization respond not to market but rather to government failures. For example, the privatization of highways often involves shifting from a system of tax-financing to a system of tolls (beneficiary-financing) to correct for the failure of the state, not of the market, to solve the collective action problem of congestion. Or, the outsourcing of state-financed service provision may be introduced in order to correct for the failure of the state to secure quality provision, because of scarce competition within the public sector. Further, markets and states can both fail, in which case, according to economic theory, nonprofits intervene (Hansmann 1980). For example, in the case of nursing homes, because of important information asymmetries due to the fact that the beneficiaries are themselves unable to assess and report on the quality of received services, states may turn to nonprofits, with the expectation that, because of nonprofits’ non-distribution constraint, which prohibits the distribution of profits to managers, such organizations will lack an incentive to exploit their beneficiaries’ vulnerability and ignorance. In any case, what matters is that, according to Heath, the merits of each case of privatization must be judged exclusively, or at least primarily in terms of efficiency. Let us call this the efficiency thesis.
In this paper, my aim is to question each one of the above theses. With regards to the asymmetry thesis, I will argue that, when it comes to privatization, the difference between core and economic functions may not be as relevant, let alone ‘essential’, as Heath suggests. Arguments against the privatization of core functions either distinctively apply to such functions alone, but they are generally untenable, or when more plausible, they would seem to extend, with similar force, to at least some economic activities as well.

As for the case-by-case thesis, I will contend that Heath’s approach to privatization problematically neglects its structural and aggregative effects—effects that give us reasons to take seriously the overall scale of privatization, beyond specific instances of it, taken serially. An instance of privatization that, if analyzed in isolation, might well be anodyne may no longer be so when seen as a further contribution to an already expansive process of privatization.

Finally, in relation to the efficiency thesis, my view is that there is little reason to think that efficiency should be the supreme value when it comes to assessing the privatization of public goods. Heath is right that appeals to distributive equality and decommodification cannot often play the anti-privatization role that its defenders would like them to play (see also Cordelli 2020, chapter 1). However, this does not exclude the possibility that other values should play an important role in justifying limits to privatization, even when further privatization would be desirable in terms of efficiency. I will argue that democratic legitimacy is one such value.

Although I disagree with Heath’s approach to the question of privatization overall, as it is summarized in the three theses above, I do not disagree with his main conclusions that, “someone who is a supporter of the welfare state might also support the privatization of certain state services in certain cases” (26) and that “examples of anodyne privatization would include […] the contracting out of public services in cases where relatively complete and enforceable contracts can be written (such as municipal garbage collection)” (59). Indeed, unlike Heath, I interpret the recent philosophical literature on privatization, including its most critical contributors, as unanimously converging on these conclusions. For example, Aviay Dorfman and Alon Harel (2013), who present inherent objections to privatization, limit their objections to the management of prisons and military functions. They make clear that their objections do not extend to welfare and economic functions, leaving open the possibility that the privatization of at least some of these functions may well be anodyne.
Similarly, Debra Satz (2018) develops a pluralistic set of standards for assessing different cases of privatization, while affirming that “the case for or against privatization almost always depends critically on context-specific details” (9). She precisely mentions the privatization of garbage collection as a generally unproblematic case. Even my own work on privatization (Cordelli 2020), which offers principle-based reasons to object to the privatization of some functions beyond putatively core ones, limits such reasons to either cases where, due to the unavoidable incompleteness of contracts or to the desirability of discretion, private actors are delegated quasi-legislative—not merely executive—decision-making powers, or to cases where privatization has reached a certain scale. I present no objection to isolated instances of “contracting out of public services in cases where relatively complete and enforceable contracts can be written” (Heath 2023, 59). Garbage collection may well be, in certain contexts, one of these instances.

I thus find Heath’s conclusion largely unobjectionable (‘largely’ because, as I have mentioned, an instance of privatization that, if analyzed in isolation, might well be anodyne may no longer be so when seen as a further contribution to an already very expansive process of privatization). This is why, instead of primarily focusing on such a conclusion, I believe it is more fruitful to focus on Heath’s overall approach to the question of privatization, as it is summarized in the three theses above.

Before I begin, a clarificatory note on the scope of my critique is in order. I suspect that much of the disagreement I have with the way Heath approaches the question of privatization derives from a more fundamental disagreement about the role of the state and its justification. Heath writes as if the state’s fundamental rationale is instrumental, and public and private institutions are just interchangeable means for the fulfillment of independently defined ends. In other words, he appears to endorse what elsewhere I have called ‘the interchangeability assumption’. The government’s main business, in this view, is solving collective action problems, and supplying goods and services efficiently, rather than instituting and constituting just power relations among free and equal persons. This means, in turn, that what states do, precisely like what businesses do, can be almost exclusively assessed according to a consequentialist, efficiency-oriented logic. Elsewhere, I have argued at length against this instrumentalist conception of the state, and against the interchangeability assumption (see Cordelli 2020, chapter 2). However, for the purpose of the present paper, I will leave these more fundamental issues aside, in the hope
that even those who do not endorse my preferred, Kantian conception of the state could agree with the substance of my critique.

II. THE ASYMMETRY THESIS

Why draw a sharp line between core and economic activities? Why think that anodyne forms of privatization only pertain to states’ ‘economic activities’, which do not constitute the coercive apparatus of the state and that, by contrast, it is “courting disaster to begin replacing elements of that apparatus with private organizations” (28)? On the one hand, Heath would seem to think that, in the case of core functions alone, there may be inherent reasons, based on “basic philosophical principles” (28) to object to privatization—reasons that are not contingent on state regulation or other contextual factors. On the other hand, he claims that:

Most obviously, because of the dangers of corruption, there is considerable merit in compensation systems that neutralize pecuniary concerns for state employees who are directly involved in the exercise of coercive power. (28)

Such systems are distinctively offered by the public sector, which can mitigate the risk of corruption by providing strong employment security, low financial incentives, as well as “strong shared values” (28). These remarks offer a more contingent argument—about both the importance and likelihood of avoiding corruption—against the privatization of core functions, which is supposed to not extend with equal force to economic activities. Both sets of arguments, I will now suggest, ultimately fail to support the claim that an analysis of privatization should draw a sharp line between core and economic activities.

What objection, stemming from ‘basic philosophical principles’ could there be that would rule out the privatization of core functions, but not the one of economic activities? Since Heath does not defend any specific objection, I will consider the main arguments already available within the philosophical literature. One such objection is that the privatization of core functions amounts to an abdication of a legitimacy-grounding responsibility on the part of the state. As put by Peter W. Singer (2003, 226), in the case of the privatization of military functions “When a government delegates out part of its role in national security through the recruitment and maintenance of armed forces, it is abdicating an essential responsibility”. Here by ‘essential’ Singer means a responsibility upon the
adequate dischargement of which the duty to obey of citizens depends, the assumption being that a state’s right to rule depends on its ability to protect its citizens’ basic rights, through the provision of policing and national defense (see Pattison 2010, 436). Such arguments, however, would not be able to rule out the privatization of core functions as we know it, because such privatizations, as Heath rightly points out (27), do not generally entail full state withdrawal but rather the contracting out or outsourcing of prison management, military, and police functions, for which the state retains primary responsibility. A state could, at least in principle, perfectly fulfill its responsibility to protect its citizens’ rights by delegating its fulfillment to a set of proxies, including regulated private actors, rather than by providing those functions directly. The objection is thus untenable.

A more plausible, principle-based objection would be that core functions cannot be legitimately performed by private actors, because the latter either lack the legitimate authority to exercise coercive powers or they cannot communicate condemnation for public wrongs ‘in the name of’ the public, which is to say, they cannot punish (27). A version of this argument is defended by Dorfman and Harel (2013). Whereas this objection arguably succeeds in ruling out the independent exercise of coercion or punishment by private agents, including all forms of vigilantism, it arguably cannot rule out cases, such as contracting out or outsourcing, where private actors are formally authorized by the state to exercise coercion or inflict punishment on its behalf, and where private actors act within the boundaries of their authorized mandate (or contract). After all, if I authorize an agent to perform an action on my behalf or to communicate a message in my name, and that agent acts within the boundaries of my authorized mandate, we can say that the agent has acted on my behalf or communicated the message in my name, and this is so even if the private agent has necessarily exercised some discretion in deciding how to act (see Cordelli 2020).

But now assume that private actors, even if formally authorized and regulated by the state, would still lack the capacity or moral standing to exercise certain powers or to make certain decisions in the name of the public, that is to say, in a genuinely representative capacity. Elsewhere I have defended a version of this position (Cordelli 2020, chapter 5). There I argue that to act as a representative of the public, an agent must be able to exclude certain reasons of a non-public nature from its deliberation and that private actors, even when acting under state authorization, fail
to meet this condition. This argument, however, does not rule out the privatization of core functions only, for the simple reason that policing and war-making are not the only functions that ought to be exercised in a representative capacity—‘in the name of all’. We may have very good reasons to want decisions on, say, the boundaries of freedom of speech within the digital public sphere to be made in the name of all, and to question on these grounds Twitter's standing to make such decisions in a representative capacity. For the same reasons we may also question Bill Gates’ capacity to set global health policy priorities in the name of all, even if such decisions do not pertain to the state’s monopoly over the exercise of coercive powers.

These cursory remarks are just meant to show that finding plausible principle-based objections to the privatization of core functions might be much more difficult than Heath suggests. If such objections can be found, they are likely to extend to some ‘economic activities’ as well.

But what about the argument from corruption? As we saw, Heath claims that the avoidance of corruption may be a strong, and perhaps sufficient, reason against the privatization of core functions. Here the concern with corruption can be understood in two different ways. First, there is the concern that the presence of pecuniary incentives will lead private actors to perform their functions in ways that undermine or corrupt the desirable or just exercise of those functions, or of the powers involved in that exercise. For example, military contractors have often been accused of fighting wars or conducting peacekeeping operations in ways that lead to either the protraction or the repetition of conflicts, rather than to their prompt resolution, because they benefit from that protraction or repetition (Pattison 2010, 446; Taylor 2015, 4). Second, there is the concern with undue influence in politics. For example, the private prison industry has been accused of lobbying for, and of being secretly involved in the drafting of, laws and policies that would increase rates of incarceration (Satz 2018). The risk here is of corrupting both the political process and the social goal of diminishing such rates.

Now, if such concerns offer a sufficient or, in any case, a very strong reason against privatizing coercive functions, why is this reason not equally strong when it comes to other forms of privatization? Consider, as an example, the privatization of healthcare in the Italian region of Lombardy—the first region in the western part of the world to be hit by Covid 19. One reason why Lombardy was found completely unprepared in its response to the pandemic, compared to other Italian regions, in spite of
being among the richest regions of Europe and of priding itself for having an excellent healthcare system, was largely a consequence of the privatization of such systems since the 1990s (see, for example, Sartor 2021). At the beginning of the pandemic, almost 50% of healthcare providers in Lombardy were privately owned and managed, although the state still acted as the main purchaser of their services. To make profits, such providers had progressively reduced the provision of preventative care and emergency services. These were the least profitable of all healthcare services, and yet the most needed not only to face pandemics but also to reduce the number of preventable illnesses across the population. Private providers had, instead, focused their investments on the provision of highly specialized, and much more lucrative, curative medicine. Although we would not generally describe the behavior of such providers as an “abuse of power” (28), the fact remains that their economic incentives corrupted, indeed perverted, the appropriate balance between preventive and curative healthcare, thereby undermining a rational and just provision of healthcare services, hence costing illnesses that could have been avoided and lives that should have been saved. Similar considerations extend to the problem of undue influence on politics. Obviously, not only entities which perform core functions can exercise such influence. Like private prisons can lobby for higher incarceration rates, so pharmaceutical companies who enjoy a monopoly or quasi-monopoly on life-saving vaccines can lobby against reforms to intellectual property that would allow the cheaper reproduction of those vaccines (see, for example, Corporate Europe Observatory 2021). Stakes for the wellbeing of citizens, in both the healthcare and the pharmaceutical cases, seem as high as in the case of abuses of power by prison managers, nor there is a reason to think that the problem of perverse incentives, and of corruption can be more easily avoided, through regulations, in the case of healthcare providers or pharmaceutical companies than in the case of prisons or military firms. Therefore, if corruption is the most obvious reason to limit the privatization of core functions, such reasons would seem to extend with equal force to at least some ‘economic activities’ as well.

Interestingly, however, Heath also claims that, at times, avoiding corruption provides a reason for privatization, not against it. In his words:

The accumulation of significant resources and economic power within the state can be, in effect, corrupting, benefiting constituencies whose interests are poorly aligned with the general public interest. (51)
Cases may arise in which “it is easier to maintain that alignment through external contracting than through internal supervisory arrangement” (51). As an illustrative example, Heath mentions the case of elected representatives from coal producing regions in the US who would likely oppose any effort at climate change mitigation. This fact, Heath says, is generally seen as a consequence of successful lobbying by the coal industry. “But now imagine”—he adds—“if state governments owned the coal companies! This would make things worse because public managers are immune from lobbying restrictions imposed on private firms” (51). Fair enough. But why, then, does not this rationale suffice to also justify the privatization of core functions, such as policing or military functions? Indeed, without great effort, one could imagine elected representatives from very red states in the US to be similarly counted on to oppose any effort at diminishing criminalization and incarceration rates, and at ending ongoing war conflicts. This could be understood as a consequence of lobbying efforts from the private prison and military industry. But now, one could say, imagine if all prisons, as well as the military security industry, were publicly owned and managed! This could make things worse for it would provide prison managers and the military with an incentive to lobby for policies that would increase criminalization rates, or for the continuation of ongoing conflicts, without being subject to the same restrictions as private firms. Why not then privatize prisons and the military industry, so as to prevent empowering constituencies whose interests are poorly aligned with the general public interest? Like the corruption-based argument against privatization seems to know no sharp line between core and economic functions, similarly the corruption-based argument for privatization knows no such line.

The point of this section has been to show that principled arguments against the privatization of core functions either distinctively apply to such functions alone, but they are generally untenable, or when more plausible, they would seem to extend, with equal force, to at least some economic activities as well. Similarly, arguments for the privatization of economic activities may extend, with equal force, to core functions. One cannot, therefore, coherently keep a sharp line between core and economic functions, supporting the privatization of the latter (at least in some cases), while ruling out the privatization of the former.
III. THE CASE-BY-CASE THESIS

I now turn to the case-by-case thesis: the idea that a normative assessment of the justifiability of privatization decisions should proceed by analyzing the distinctive benefits and costs, in terms of efficiency, involved in particular instances of privatization, taken separately. Heath's case-by-case approach has the advantage of clearly showing how and why the benefits (for example, decongestion) and thus the reasons that may prima facie justify the privatization of certain goods, say, highways (that is, moving from a tax-financed system to a toll system) may be very different from the benefits (for example, increased competition), and thus the reasons that may prima facie justify the privatization of other goods, say, public service provision (for example, the outsourcing of healthcare or welfare provision). Similarly, the costs involved in some types of privatization (for example, the problem of contract incompleteness in cases of outsourcing) may be absent in other types of privatization (for example, the selling—transfer of ownership—of a public industry).

The problem with a case-by-case analysis, however, is that it both neglects and obscures the aggregative, structural, and dynamic effects of privatization, especially on (i) the overall balance of power between the public and the private, states and private corporations; and (ii) the relationship between citizens and their state.

III.I. The Balance of Power between the Public and the Private

The overall scale of privatization, above and beyond the particular kind of privatized functions and the specific types of privatization, affects both the level of dependency of the state on the private sector and its capacity to regulate and control the later.

Dependency arises because the more a government privatizes, the more it loses in-house capacities and know-how to perform a variety of functions directly, the more it becomes dependent on the private sector for the performance of essential tasks. In turn, the higher its dependence on the private sector, the more the state will be vulnerable to, and powerless in front of, pressures from such a sector—a sector populated by actors with a vast amount of economic resources and, in a context of international competition, with the power to threaten to bring their resources elsewhere. In the words of Henry Farrell (2018):

As more aspects of the economy are privatized, it becomes easier for actors who might benefit from privatization to press the state to make
further concessions. As states become more reliant on the private sector for information and resources, they become more inclined to acquiesce to the demands of private actors. (175)

The problem of dependency is a problem of scale, for whereas the privatization of some, say, military functions or IT functions may not suffice to put business in a privileged position vis-à-vis the state, the privatization of most or all military and IT functions will. This is also because, the wider the loss of capacity and know-how on the part of the public sector, the more difficult it will become for the state to initiate a process of insourcing to avoid undue pressures. The problem of dependency is further worsened by the likelihood of a ‘brain drain’ from government to the private sector, which can generally afford to pay higher salaries (Freeman and Minow 2009; Verkuil 2007; Michaels 2017).

In sum, privatization can generate a form of structural dependency, which in turn changes the overall balance of power between the state and market actors, giving rise to a situation where the latter have the incentives, resources, and power to pressure and, ultimately to dominate, the former. Because the problem of dependency is conditional on the overall scale of privatization, it remains invisible from the perspective of a case-by-case approach.

One could, however, respond that the problem of dependency can in large part be resolved by means of regulation. But this response risks being circular, because the state’s effective capacity to keep private actors under appropriate control and accountability standards itself depends on its level of dependence on the private sector—sector that can use its resources to impede needed regulations. Even if we leave this problem aside, there are further reasons to think that the scale of privatization can undermine a state’s regulatory capacity. This is because, as Heath himself acknowledges (47), if a government privatizes because it lacks the expertise to solve a certain problem, then, it will also likely lack the capacity to design contracts that specify its needs with a level of detail sufficient to protect itself from exploitation by private contractors. To which we may add that, if government lacks the capacity to directly perform certain functions or to do so efficiently, it will also likely lack sufficient capacity to coordinate, plan, oversee and regulate those to whom those functions are delegated, and to do so efficiently (Freeman and Minow 2009). This problem is made more acute by the transnational nature and international mobility of many private actors, which makes their tracking and
monitoring particularly difficult. Take, as an example, the progressive expansion of contracting in the U.S. military during the first decade of the 21st century. The Department of Defense’s contracting budget increased by more than 100% between 2000 and 2005. This expansion led to a situation in which public officials did not even know, and were not able to easily find out, how many private contractors were working for the military (Minow 2009, 16). Monitoring was rendered particularly difficult by the transnational and mobile nature of private security firms.

This *problem of control* is further intensified by both the likelihood of brain drain from the public to the private, and by the fact that, insofar as government agencies often privatize to save costs, they are then reasonably reticent to hire new personnel to supervise their private contractors. For example, while federal spending in the USA exponentially increased between 1960 and the first decade of the 21st century, much of which was directed to finance newly contracted services, the number of civilian workers remained pretty much unchanged (Dilulio 2015). Who was then supposed to monitor and supervise the also exponentially higher number of proxies (about 40% of the workforce employed by the U.S. government) who were required to manage that spending? This led to a situation in which, as Paul Verkuil (2009, 313) puts it, “if government does not have adequate personnel to oversee its outsourcing, it does not have adequate personnel to read the reports on outsourcing submitted by its private overseers.” In other words, the more government outsources, the less capacity it is likely to retain to gather basic information about performance, costs, and outcomes, and thus to choose competent contractors, as well as to enforce contractual terms.

Finally, as law scholar Jon Michaels (2017, 131) explains, privatization often achieves its purposes by “swapping out heavily regulated and duty-bound civil servants for less regulated (and therefore more pliable) private contractors”. Indeed, many of the standards of accountability that apply to public actors do not extend to private contractors. Therefore, the more expansive the scope of privatized functions, the lower the degree of control that a democratic state, and thus also its citizens, retain on the performance of those functions.

The problem of control, like the one of dependency, is a problem of scale. It is a consequence not of particular instances of privatization, taken serially, but of the aggregate effects of the systematic use of outsourcing on the capacity of the administrative state to retain adequate
control over the private sector, when such a sector is hired to act as its agent. A case-by-case approach risks obscuring this problem as well.

III.II. The Relationship Between Citizens and their State

But what about the effects of privatization on citizens’ relation to the state? Heath argues that “citizens formulate their opinions about the qualities of the state on the basis of their interactions with state-level bureaucrats” and that “to the extent that they develop a negative attitude, they tend to apply it to all its activities, optional or not” (61). Insofar as public sector organizations lack the right incentives to provide polite and responsive costumer service, Heath concludes that states have reasons to avoid providing specific, optional services, such as running airlines or hotels. Perhaps. But Heath does not consider the overall effects of privatization on citizens’ attitudes towards the state. There are, indeed, important reasons to believe that the privatization of public services, when expansive in scope, has significant effects on citizens’ civic vigilance and, more broadly, on their interest in politics.

One reason, and a direct and foreseeable consequence of the loss of control analyzed above, is that privatization makes abuses less detectable. As Jody Freeman and Martha Minow (2009) observe:

Outsourcing impairs the visibility necessary to check for such abuses [like waste and fraud], because private companies control info about cost, performance, and other vital data that otherwise would be open to review by government agencies. (5)

Now, the harder it becomes for citizens to find out what happens, the more they should be vigilant, but the less inclined to be vigilant they will be. For if citizens cannot see that something is wrong, they will take themselves to have little reason for vigilance. Further, the harder it is for them to find information about abuses, the more demanding their exercise of vigilance will become, the less inclined to be vigilant they will be.

But there is more. The privatization of public services, even when limited to their provision, occludes the role of government behind a myriad of market and charitable actors through which the government comes to provide needed benefits to its citizens. Privatization thus directly contributes to what Suzanne Mettler (2011) has called the ‘submerged state’. “The policies of the submerged state”—Mettler explains—“remain largely invisible […] Even when people stare directly at these policies, many perceive
only a freely functioning market system at work” (5). This in turn “leav[es] citizens unaware of how power operates, unable to form meaningful opinions, and incapable, therefore of voicing their views accordingly” (5).

Insofar as, in their daily lives, citizens experience government mainly through the provision of goods and services, when people do not perceive their own government as the main provider of the benefits they need, they see little reason to care about their government and thus to actively participate in politics. It follows that, in contexts where ‘the face’ of government is largely privatized, citizens' interest in politics, including the urge to ensure their government’s accountability through continuous vigilance, tends to diminish (as the empirical literature on the submerged state confirms), and civic apathy to grow. Generalized across a polity, therefore, pervasive privatization “contribute[s] to civic disengagement, including ignorance of public affairs, disenchantment with government and political apathy” (Norris 2000, 309. See also Cordelli 2020, chapter 4). Privatization thus generates a problem of vigilance, beyond those of dependency and control. Like the latter, also the former is not a problem that can be captured by a ‘case-by-case’ analysis of privatization, for the outsourcing of some specific healthcare services to private providers here and there or the existence of a few private schools or childcare centers as supplements to public providers, may not create any problem of civic apathy. But the systematic outsourcing of those very same public services may.

The point of this section has been to argue that instances of privatization cannot be assessed ‘case-by-case’, without taking into account the aggregative effects of privatizations, as well as the scale of privatization already in place in a given society. A same decision to privatize healthcare services may have different implications for the overall balance of power between the public and the private sector, as well as for the relation of citizens with their state, in different societies, depending on what else and how much has been already privatized in those societies. This interestingly also means that cases of privatization that may count as perfectly anodyne, if analyzed in isolation, may become more problematic if seen as contributions to the further privatization of a system of provision that is already largely privatized. A structural or systemic analysis of privatization is thus needed, above and beyond a ‘case-by-case’ one.

1 For an application of this argument to developing countries, for example, see Dambisa Moyo (2009).
IV. Efficiency Thesis

If my, admittedly cursory, analysis of the three problems of dependency, control and vigilance is sound, then privatization, especially when expansive, would seem to generate a serious problem of democratic legitimacy. Avoiding such a problem, in turn, provides us with strong reasons to limit the scale of privatization, even when imposing these limits would come with some efficiency costs.

The problem of democratic legitimacy I have in mind has three dimensions. The first dimension follows from the problem of dependency and amounts to the domination of the state by private actors. Privatization provides private actors with incentives to pressure the state to enact policies that will continue the privatization process. As we saw, the more extensive the privatization of public functions, the more dependent the state on the private sector, the more powerless it becomes to reject those pressures, or to start processes of insourcing to reduce them. The result is both the risk of domination of the democratic process by private interests and the violation of a basic principle of political equality—some parties (that is, market actors) enjoying much more political power and opportunities for political influence than others. Both concerns, note, directly threaten the legitimacy of the democratic process, regardless of whether they lead to more or less efficient outcomes.

The second dimension follows from the problem of control and amounts to the domination of citizens by a privatized state. Here the problem is that if, (i) due to the pervasive problem of contract incompleteness, private contractors are often left with wide degrees of discretion on how to interpret the contract and to perform relevant functions, and if, as we saw, (ii) the more a state privatizes, the less likely it is to retain the effective capacity to exercise appropriate forms of control, monitoring and supervision over private contractors, then, the obvious result is that (iii) citizens become subject to unaccountable exercises of discretionary powers, that is to say, subject to the arbitrary discretion of private state contractors. Such discretion may concern decisions such as: the setting of eligibility criteria for the allocation of unemployment benefits or welfare services; the prioritization of claims to certain kinds of medical treatment over others; determinations concerning the kind of infringements for which inmates can be sanctioned within a prison; and so on and so forth (see Cordelli 2020). The problem with cases in which such discretion is exercised arbitrarily, because private state contractors remain insufficiently monitored or unaccountable, is not that the democratic state is
dominated by private interests but rather that citizens are dominated by an internally privatized state that rules their lives arbitrarily, through its contractors. Once again, this concern is independent of whether private actors exercise their discretion in ways that benefit or rather undermine overall efficiency.

The third dimension follows from all the three previously analyzed problems—dependence, control, and vigilance—and amounts to the abdication of the necessary conditions for democratic self-government, through their progressive erosion (see Cordelli 2020, chapter 4). In a society where: (i) the political process is systematically dominated by private interests and thus the basic principle of political equality is violated; (ii) citizens lack efficacious control, through their institutions, over those (private) agents who are left with the responsibility to manage public funds or to provide essential public services to them; and (iii) the risk of civic apathy and indifference is widespread, then the basic conditions of democratic self-government are compromised. Yet such a society is, as we saw, a foreseeable consequence of the systematic privatization of public goods, functions, and services, precisely because of the problem of dependence, control, and vigilance that such privatization structurally produces. I say ‘structurally’ for many of the causes of such problems do not pertain to the subjective inclinations or corrupt intents of particular agents (see Young 2011). Rather, they derive from objective conditions, such as the presence of specific economic incentives; the overall balance of power between the public and the private; the transnational nature of many private actors, as well as the existence of international competition; the problem of brain drain, which follows from the fact that market actors can pay higher salaries; and so on and so forth.

Jointly taken, the three concerns just analyzed—the domination of the political process by private agents; the domination of citizens by a privatized state; and the erosion of the basic conditions of democratic self-government—amount to a very serious deficit of democratic legitimacy. We may call this ‘the problem of democratic legitimacy’. The prevention of this problem provides strong reasons to limit the overall scale of privatization, even when more privatization would mean more efficiency.

Why, one may ask, should legitimacy have priority over efficiency? I cannot provide an exhaustive answer to this question here. I will limit myself to offer a hypothetical case that, albeit far-fetched, provides intuitive support to the priority of legitimacy (see also Stilz 2016). Imagine Germany tomorrow conquered Italy but only decided to take over the
management of the Italian welfare state, promising higher efficiency and quality of service, while leaving other domains of state action untouched. Imagine Germany delivered on its promise. Italians would still have a complaint, and I would say a rather strong one, against the German takeover, and this is so even if we assume such a takeover could happen without the use of violent means. The complaint would be grounded on the value of collective self-determination, and on the lack of legitimacy of the German government vis-à-vis Italians. Italian citizens could rightfully complain that, despite its efficiency gains, the takeover would deprive them of the right to collectively decide, as equals, on important decisions concerning their society, to control those decisions by holding accountable those empowered to make them, and would also subject them to a dominating, because alien and unaccountable, will.

Similarly, even if the systematic privatization of ‘economic activities’ could lead to more efficient outcomes, citizens may still reasonably complain that, by giving rise to the legitimacy problem, such privatizations would subject them to an arbitrary form of ruling and would deprive them of the basic conditions of collective self-government.

In sum, once we leave behind a case-by-case approach and focus on the aggregative and structural effects of privatization, legitimacy concerns enter the scene. Once such concerns are considered, it becomes clear that there may be very good reasons to limit the overall scale of privatization, even in cases where more privatization, by assumption, would come with efficiency gains. Efficiency, it seems, is not the only, or even the primary value in light of which the ‘merits’ of privatization decisions should be assessed. Further, anodyne forms of privatizations may become more problematic, when seen as further contributions to an already expansive process of privatization.

V. Conclusion

In this paper I have argued, contrary to Heath, that a philosophical assessment of the justifiability of privatization decisions (i) need not start by drawing a sharp line between core functions and economic activities; (ii) should not exclusively proceed on a case-by-case basis, thereby neglecting the aggregative and systemic effects of privatization on the overall balance of power between the public and the private, and the relationship between citizens and their state; and (iii) should assess the (de)merits of privatization not only in terms of efficiency, but also in light of other values such as legitimacy. I have further suggested that the scale of
privatization matters beyond the kind of privatized functions and the modalities of privatization. Instances of privatization that, if taken in isolation, may well be anodyne, become more problematic, normatively speaking, when seen as further contributions to an already very expansive process of privatization.

REFERENCES


Chiara Cordelli is Professor of Political Science at the University of Chicago and the author of *The Privatized State* (Princeton University Press, 2020).
Contact e-mail: <cordelli@uchicago.edu>
Repoliticizing Privatization

SAVRIËL DILLINGH
Erasmus University Rotterdam

Abstract: According to Joseph Heath, privatizations should be judged on a case-by-case basis with appeal to the Pareto criterion. This approach, or so I argue, amounts to a depoliticization of privatization. While Heath’s approach is effective and at times illuminating, I show that a consistent application of his methodology is self-defeating in that it eventually requires a politicization of privatization. With appeal to transaction cost theory, I show there are social costs associated with affirming the competitive pressures of the market. Subsequently, I argue that while private actors may, according to Heath, pursue efficiency with appeal to an adversarial morality, state-owned enterprises (SOEs) are much more constrained in how they may achieve such gains. Due to the pressures of liberal neutrality, SOEs may chase efficiency only without setting actors back. Conversely, the private sector's potential for success is predicated on its ability to compete for Kaldor-Hicks efficiency, which specifically allows for win-lose interactions. While SOEs are often no more able than the private sector to achieve Pareto optima, this discrepancy makes it so that SOEs produce gains that are ex-ante lower but more equal, whereas the private sector produces gains that are ex-ante higher but more unequal. Thus, the social cost of affirming the competitive pressures of the market is ex-ante inequality. If this premise is accepted, there is no way to avoid the conclusion that a consistent case-by-case approach requires a structural, political view on how privatizations affect the state’s ex-post Pareto-enhancing abilities.

Keywords: Pareto, Kaldor-Hicks, market failures, SOE, state-firm analogy, transaction costs, cooperation, competition, liberal neutrality, inequality

JEL Classification: H4, H6, L2, L3, P16
I. INTRODUCTION

While the Dutch railways were slowly being privatized between 1995 and 2002, ministers overseeing the process had to contend with a persistent wave of public discontent. The privatization project had been plagued by severe teething troubles, resulting in overfull trains, cancellations, and personnel shortages, as well as a palpable sense of patriotic loss. Something, so the public felt, was being taken from them. The preceding decades had already been characterized by widespread privatizations of many traditionally state-governed industries. Yet where an earlier governmental consensus touted the virtues of the market in all contexts, modern critics had grown steadily dourer. They pointed to the staggering failures of several privatization efforts, at least in part due to—or so the argument went and still goes—a colonization of the public by the private.

Naturally, the subject has not left political philosophers cold, either. In more recent years, calls for a reevaluation of the supposed boons of privatization have seen a marked resurgence. Many of the most popular criticisms adopt a distinctly deontological hue (Cordelli 2020; 2021).¹

Joseph Heath’s article is a nuanced, thoughtful counteroffensive to the counteroffensive. It is also simultaneously an extension to and combination of his work on the market failure approach (Heath 2023b) and his ideas about public governance (Heath 2022). On Heath’s view, there is no knockdown argument against privatization as such. Rather, any given instance of (de-)privatization should be judged on a case-by-case basis against the value of Pareto efficiency. “This is a view”, so says Heath, “commonly held by economists, and much less commonly held by philosophers and political theorists” (2023a, 60). Values like equality, justice, or other “basic principles of political philosophy” (30) are often irrelevant to a public manager’s generally non-ideological “pragmatic managerial decision-making that merely adjusts the inevitable boundary between public and private” (31). On my view, Heath’s account thus amounts to a depoliticization of privatization.

Yet depoliticization need not be a dirty word. In fact, Heath’s depoliticized account of privatization is immensely attractive! Its considerable strength derives not only from its intricate systematization of the various forms privatization may take, but also from its common-sensical formulation. That is, if one accepts—like many economists do—that “state involvement in the economy should be guided primarily by the norm of

¹ A significant exception is Satz (2019).
[Pareto] efficiency” (60), then it would certainly be prudent to assess whether this norm has been satisfied on a case-by-case basis. Indeed, this rather modest conclusion is difficult to argue with.

Of course, many philosophers and political theorists alike reject this type of Paretian view wholesale. At least part of the reason for that is the growing body of economic data—as well as robust philosophical argumentation—suggesting that a healthy skepticism vis-à-vis whether Pareto efficiency should be considered a primary, and not just an instrumental, moral value is warranted (Piketty 2021; Vrousalis 2023). This is particularly poignant in the case of Pareto wealth inequality—astronomical (and rising) inequality that is, nonetheless, perfectly Pareto efficient (Jones 2015; Robeyns forthcoming). Insights like these inform the view that, although the state should by no means completely disregard efficiency-considerations, there are other, oftentimes much more important values involved in statecraft—most particularly those pertaining to the maintenance of a cooperative social ethos (Cohen 2008; Furendal and O’Neill 2022).

Of course, there are very strong reasons for generating the desired ethos in the most cost-effective manner, but this is far from positing that Pareto efficiency should be the state’s lodestar value. Thus, I think Heath’s ongoing attempt to articulate the state—even just its economic functions—as an institution chiefly geared towards efficiency gains is largely beside the point.

Yet because his position is commonly held by many a chief economic advisor and policy maker, it is worth meeting Heath head on, on his home turf. This entails adopting not only the Paretian ‘Market Failures Approach’ (Heath 2014), but also the relatively inoffensive conclusion that follows from it with regards to privatization. By depoliticizing the practice, Heath seems to sidestep several important considerations that make the question of privatization much more amenable to ‘basic principles’ of political philosophy. However, or so I argue, taking seriously his own approach eventually obliges him to also take seriously the question of structural power between private and public actors regardless. Crucially, my account shows that this does not require importing deontological considerations, nor adopting a socialist political morality: a simple appeal to

---

2 Incidentally, this is often where the core of modern egalitarian resistance to privatization seems to stem from: not from an appeal to an ideological “distaste for the commodity form” (Heath 2023a, 35) nor “promoting greater equality” (60) in the myopically distributive sense. Although Heath may be forgiven for thinking so, because a majority of egalitarians seem curiously unaware of this themselves.

efficiency suffices. In brief, even depoliticizing privatization cannot shield Heath from its immanently political nature.

The argument will unfold as follows. Section II will expand on and adopt Heath’s usage of the ‘transaction-cost approach’, arguing that while Heath incisively presumes there are social costs associated with not privatizing, there must also be similarly social costs associated with, in fact, privatizing. Section III will go on to establish that there is a marked distinction between how privately-operated enterprises and state-operated enterprises are morally authorized to approach the question of efficiency maximization. Section IV consolidates the previous two sections by demonstrating that this distinction points towards the idea that social costs associated with privatization are best expressed as ex-ante inequality. Section V shows how this conclusion forces Heath’s methodology to include political sentiments so as to further a view on the state’s ability to turn ex-ante Kaldor-Hicks gains into ex-post Pareto gains. Section VI concludes.

II. THE TRANSACTION-COST APPROACH TO THE STATE
To set the stage, Heath invokes a particular economic parlance in his argument. In meeting him on his own territory, this parlance will prove to be extremely productive. To illustrate that setting the boundary between public and private is chiefly an efficiency consideration, Heath appeals to the ‘make or buy’ decision which “has figured centrally in the transaction-cost theory of the firm” (2023a, 30). The idea is as follows. Before the early 20th century, economic theory tended to treat firms as essentially black boxes; this changed with Ronald Coase’s landmark 1937 paper “The Nature of the Firm”. Coase asked a brilliantly simple question: If the market is so efficient, why are there firms? Indeed, why do these black boxes, strangely authoritative in nature, even exist when the market is able to allocate resources perfectly well through competitive means?

The answer Coase landed on was equal parts brilliant and simple: if there are firms, that must mean there are certain hitherto undiscovered costs associated with using the competitive price mechanism. Firms, after all, are characterized by “the supersession of the price mechanism” (389), after which the “entrepreneur-co-ordinator, who directs production” (388) takes over. This regime, Coase notes provocatively, is not at all unlike “what is normally called economic planning” (388).

---

4 This insight is taken to entail very productive, albeit contested, conclusions in Phillips and Rozworski (2019).
Of these costs associated with using the price mechanism—transaction costs—Coase brought to light three. And although modern transaction cost economists, building on Coase’s insight, have meticulously specified and added to the concept, just one will serve to illustrate Heath’s point. Bargaining costs, or the “costs of negotiating and concluding a separate contract for each exchange transaction which takes place on a market” (390–391) were, according to Coase, able to severely hamper the price mechanism’s efficacy. Indeed, metaphorically bargaining with every fisherman for every single fish caught can obviously be nigh-eliminated by drawing up long-term, incomplete contracts.

By dint of the above, it is not difficult to see that in many cases, it is simply cheaper to ‘make’ a product in-house by hiring long-term employees than it is to suffer the costs associated with having to ‘buy’ a market contract for that same product. And, in “strikingly similar” (2023a, 30) fashion, or so Heath argues, the decision whether a service should be privatized or kept ‘in-house’ falls mainly to a comparing “the state’s ‘agency costs’ [...] to its ‘contracting costs’” (31).

Perhaps a knee-jerk reaction to this argument would be to accuse Heath of surreptitiously invoking the state/firm analogy—a device nowadays often considered a non-starter for various reasons (Frega 2020). This would be a mistake; Heath is doing no such thing. He is merely arguing that the state’s economic functions appeal to the same considerations as the firm’s economic functions. This is a far cry from appealing to an analogy that takes state and firm to be similar in structural, philosophically salient ways. Indeed, Heath takes special care to separate the state’s core functions from its economic activities (2023a, 26).

Yet Heath takes his transaction cost approach to the state even further. Heath argues that there are, like transaction costs that are associated with using the price mechanism, also “some hidden costs associated with state involvement” (60) that do not arise in the private sector. That is, in furnishing a counterargument to the egalitarian rationale holding that inefficiency is “a cost worth bearing for achieving the more egalitarian outcomes” (60), Heath supposes that bad state performance in optionally public environments—environments in which public provision is not a requirement—often causes citizens to develop structurally negative attitudes to state involvement in general. In turn, so says Heath, this can easily evolve into “easy fodder for right-wing politicians with an ideological hostility to government activity in all domains” (61). In brief, what Heath

---

5 See, famously, Oliver E. Williamson (1975).
is demonstrating, in astute fashion, is that there are also costs (or risks of costs) associated with not using the price mechanism. These costs, however, are of a social caste.

It must at this point be noted that this observation may appear more profound than it really is. Indeed, a very similar phenomenon had been hypothesized some decades ago by Henry Hansmann (1996). Hansmann deduced certain “costs of ownership” on the level of corporate governance which feature most prominently in democratic firms, pertaining to the costs of collective decision making (45); call these cooperative costs. The general idea is that decision-making does not take much time, nor deliberation, for firms that appeal to shareholder primacy. Shareholders simply tend to want the same thing: to maximize shareholder value. In turn, this drastically lowers collective decision-making costs in comparison to, for instance, worker cooperatives where different workers may have wildly differing interests and preferences depending on their position in the collective, which department they work for, which colleagues they personally admire, and their idiosyncratic ambitions.\(^6\) Insofar as, on Heath’s diction, the state’s economic functions resemble the firm’s economic functions, one can see how his process may be construed as a ‘cooperative cost’.

Yet it takes some analytical finesse to see Hansmann’s discovery as a particularly social cost. Indeed, an ostensibly more unambiguous example is the converse instance that socialists of various stripes have long perceived. According to these scholars, it is using the price mechanism that in fact comes paired not just with transaction costs, but with social costs, as well (Arnold 2020; Maguire 2022; Vrousalis 2012). Call these competitive social costs. This extremely large literature on how the market may inspire greed, forces us to instrumentalize each other, or excludes certain individuals (cf. Frye 2023) is, of course, highly contested. More recently, however, less controversial literature has in a similar vein and relevant to our purposes argued that continual governmental outsourcing results in

\(^6\) Abraham Singer, in contrast, argues that “the time and energy, or even the money, spent on democratic decision-making might not be deal-breakers”, and that “workers might be willing to bear these costs for the ability to democratically control the workplace” (2019, 151). The real problem, so he posits, are the costs incurred by poor decisions—qua efficiency-considerations—that may follow from democratic control. Yet this view is certainly not uncontested, either. Felix Gerlsbeck and Lisa Herzog (2020), for instance, have argued that epistemic arguments surrounding democratic costs traditionally used against worker ownership, should in fact count in favor of it. On their view, workplace democracy can “effectively harness latent information distributed throughout a company, allow for multi-perspectival decision-making, and facilitate adaptation and reflexivity” (329).
a decimated, infantilized public sector (Mazzucato and Collington 2023). Somewhat surprisingly, various libertarian writers have also run with this originally socialist premise, yet argue that the social externalities of the market are, in fact, broadly positive (Tomasi 2012; Brennan and Jaworski 2015). Of course, there are also positive social externalities associated with public provision; where failing public institutions may erode it, they may also facilitate trust in a way the private sector cannot when they do work (Faulkner 2018). Heath is, of course, duly aware of this (2023a, 61). The upshot, however, is that various theorists agree that, in addition to efficiency costs associated with using the price mechanism, there are social costs (or benefits) associated with affirming the competitive pressures of the market.

All this goes to say that Heath’s appeal to transaction cost theory points to a social phenomenon looming over the question of privatization in general—a phenomenon that is, unfortunately, not easily perceived when utilizing a strictly case-by-case methodology. And while it may seem somewhat uneven that Heath temporarily departs from evaluating each separate case of privatization “on its merits” (41) to further a structural (and compounding) argument against public provision, his insight does provide us with a method. In brief, like Heath argues that there are cooperative social costs involved in matters of public provision, it is clear that privatization may incur similar competitive social costs, which make singular instances of privatization even trickier to assess. In the next section, I will argue that a conception of competitive social costs does, in fact, not need to depart from Heath’s approach by appealing to controversial socialist assumptions. Indeed, these social costs can be cogently expressed in terms of efficiency.

III. What Kind of Efficiency?
To express competitive social costs in terms of efficiency, however, it is imperative we specify what kind of efficiency we are talking about, as well as determine how the economic functions of both public and private sector relate to those conceptions. While I agree with Heath that the firm and the state’s economic functions are comparable in a relevant philosophical sense vis-à-vis privatization, without falling prey to the firm/state analogy, it bears keeping in mind that many modern theorists reject this analogy for an extremely pertinent reason. That is, firms, unlike states, draw their legitimacy from their “functional role” (Frega 2020, 19). According to Abraham Singer, what characterizes the firm in normative terms is “the
competitive market environment in which the corporation operates, which presents itself as crucial given the *efficiency rationale* we have ascribed to firms" (2019, 140, emphasis mine). In essence, Singer's insight is the philosophical operationalization of Coase's economic insight: if firms are not more efficient than the market, why would there be firms at all? And given that there would be no firm if it were not more efficient than the market, it follows that efficiency must always remain the firm's *raison d'être*—lest it ceases to exist. Crucially for our purposes, it is important to note that the state lacks such a demanding efficiency rationale. Indeed, the egalitarian view that Heath resists suggests that the state is sometimes *required* to be inefficient in order to achieve more equality. As I will show, it is this crucial difference between firm and state that results in both institutions differing severely in their 'efficiency mandates'—that is, the manner in which they are morally permitted to pursue efficiency. This fact, in turn, will eventually make it difficult for Heath's methodology to entirely circumvent more nakedly politicized (and egalitarian) considerations. More specifically, in this section I show that where privately-owned enterprises may (indeed, must) pursue Pareto efficiency indirectly, state-owned enterprises are permitted to do so only directly.

**III.I. Competition, Markets, and Pareto Efficiency**

The genius of the market—indeed, why it works like it does—is that it repurposes the otherwise deleterious free rider incentive. The incentive unfortunately tends to eat away at gains from cooperative endeavors; more specifically, it tends to generate compliance problems. Readers of this article should be intimately familiar with the phenomenon, since most student group assignments suffer from it. Teachers rarely do not suspect this one particular student (you know the one) of being a free rider that has probably contributed to the project much, much less than their peers. ‘Probably’ is the operative word, here, since the fact that we can never know for sure makes it so that “patterned arrangements are often feasible only where there is close to perfect observability of contribution as well as a mechanism in place to punish free-riders” (Heath 2023b, 55). It follows that the more actors enter a cooperative scheme, the more deleterious the free rider incentive becomes, and the less feasible the cooperative scheme.
Sometimes, however, competitive schemes are able to put the free rider incentive to work instead. By offering ‘the best’ participant some type of substantial reward, compliance problems can be circumvented. Competitors no longer have to care about the team, or a collective purpose, of the cooperative scheme; they only have to want to win. Importantly, however, winners now have to win at the expense of other participants. In essence, competitors have to “defect rather than cooperate” (Heath 2014, 96). The gains that can be made through the competitive scheme often far outweigh the gains that could have been made cooperatively, however.

The market is precisely such a competitive scheme, and one Heath has analyzed extensively. According to Heath,

> there is nothing magical about the ability of markets to transform private vices into public virtues. This sort of laundering is a general feature of all competitively structured social interactions. And like all other forms of competition, market competition must be governed by a set of rules, restricting the range of strategies that individuals may employ, in order to ensure that it remains healthy. (2014, 101, emphasis mine)

If properly restricted and regulated, market competition through profit maximization tends to generate Pareto efficiency; per the fundamental theorems of welfare economics, so says the familiar formula, results making “at least one person better off without making anyone worse off” (2023b, 31) are entailed by the disciplining nature of perfect competition (Arrow 1973).

It is at this point crucial to note that a new set of competitive compliance problems may arise in any given competitive scheme, including the market. More specifically, this particularly competitive form of a cooperative structure entails it is liable to devolve into a ‘race to the bottom’. The global financial crisis of ’08/’09 should now immediately come to mind. But, of course, we must note that this liability is the very reason that the market works so well when it does work. Indeed, competitive attitudes can bring about a “much higher level of performance than anyone would have thought possible […] it may produce outcomes far superior to what mere compliance could have achieved” (Heath 2023b, 58). In summary, because competitors do no longer have to be concerned with

---

7 I follow Hussain (2020) in considering competition to be just another cooperative scheme, yet one that appeals to a set of rules that structure competitive interactions.
the team or a collective purpose, they are “merely obliged to respect certain outside boundaries in the pursuit of their private interests” (Heath 2023b, 66, emphasis mine). What these ‘outside boundaries’ consist in is expertly defined by Heath’s Market Failures Approach (2014); but, unfortunately, the very structure of a competitive scheme “precludes this sort of high-mindedness” (98), and so often the outside boundaries will have to be forcefully maintained. As we have seen time and time again, a failure to maintain healthy market competition—to restrict the range of strategies allowed—sees the inherent liability of competitive schemes result in market failures that are utterly ruinous to society at large. But if everything works, it works really well.

All this goes to say that Pareto efficiency is essentially an accident. It is, as Heath has it, “necessarily a byproduct of the competitive activity” (98). This means that firms do not, in fact, strive for Pareto efficient outcomes. Indeed, they should not, because the very idea of Pareto efficiency entails a focus on winning, which in the market context amounts to maximizing profits. Call this this phenomenon the ‘market efficiency mandate’.

It should be no surprise, then, that actual Pareto optima are few and far between. The idealized circumstances required for perfect competition are rarely, if ever, met. But this is not the point of Heath’s Market Failures Approach. Heath’s argument is simply that Pareto efficiency should be the market’s normative standard. In reality, markets much more often produce Kaldor-Hicks efficient outcomes. The Kaldor-Hicks norm denotes an interaction that is simply “aggregate-welfare-enhancing” (Moriarty 2020, 118). That means that while the Pareto standard is restricted to generating win-win situations, Kaldor-Hicks may produce win-lose situations as long as the winner’s gains are large enough to compensate the loser’s losses. It is, so to speak, unconcerned with if gains made are distributed ex-post, as long as they could be. The Pareto standard, in contrast, can be expressed as a pre-distributive, relational principle: gains may be made only such that there are no losers, and so that winners cannot win at the expense of others. This is, of course, why only Pareto efficiency is relatively uncontroversially considered to be not just an economic, but also a moral principle. Kaldor-Hicks efficiency specified as such is much, much more difficult to defend on moral grounds.8

---

8 According to Singer, “to be relatively indifferent to efficiency is to be relatively indifferent to human welfare” (2019, 27). It is extremely difficult to imagine that he intends to refer to Kaldor-Hicks efficiency here.
III.II. Competition, Schmarkets, and Pareto
Governments have, of course, long realized the market’s ameliorative qualities. Indeed, the past century is perhaps best characterized as the period in which the state had “discovered that an institution possessing a monopoly of force could also be an important economic actor” (Heath 2023a, 27). While its economic endeavors have not always been successful, there is something significant about the behavior of state-owned enterprises (SOE), or so I will show, that recontextualize the question of privatization when analyzed over lines of efficiency.

Since the state lacks the firm’s efficiency rationale, it is able to adopt multiple—sometimes even conflicting—values. Indeed, this is why SOEs are often derided for “the possibility that in conducting business, government-owned or -controlled entities may [...] substitute political ambitions instead of or in addition to profit-making” (Malkawi 2019, 23). Yet it is precisely this ability that allows government to “accomplish certain tasks that no private actor seemed prepared to take on” (Heath 2023a, 27). As is well known, the state’s economic endeavors are not necessarily motivated by the promise of profit, but often include a range of noncommercial objectives motivated by “the presence of significant positive externalities” (36), as in the case of education, or because the state is in various circumstances “not vulnerable to adverse selection” (36) or moral hazard (see also Horne and Heath 2022). Heath is—and so am I, in staying with his logic—inclined to explain this away as the state correcting market failures. And while this may be the case, it is important to in addition note that it is able to correct those market failures in large part because it lacks the firm’s efficiency rationale.

As a result, the state is able to erect ‘schmarkets’—competitive environments that simulate market competition where it would, in actuality, be impossible to sustain a real market due to a crippling presence of, for instance, information asymmetries or negative externalities. Schmarkets may even see the state have several SOEs compete with each other, such as is often the case in banking and finance, and—in many West-European countries—in transportation; or the state may allow private actors to compete with each other in a heavily subsidized schmarket, such as in several countries’ telecommunications sector. The point is that prior to erecting a schmarket, the state has already proceeded to make several normative and structural decisions concerning what exactly this artificial market should look like.
Now, the distinction between the state correcting a market failure and erecting a schmarket might seem academic in the pejorative sense. But it is significant. One can see the distinction’s relevance when one considers the market’s historical emergence as a state-induced structure (Graeber 2011). More specifically, markets are not natural phenomena; they are designed by states (Murphy and Nagel 2004; Roth 2015). To construe the state as only ever being able to correct market failure, even to the extent that meticulously subsidized, otherwise non-existent markets are seen as market failures mended, implies that markets are a form of pre-state institution. This is both historically and analytically incorrect. In fact, the very fact that a schmarket is erected where a relatively self-sustaining market would never survive, implies the very opposite. That is, when a state has a stake in a first-order objective it may, in order to achieve it, select the market mechanism as a second-order tool. Per Heath’s logic, then, this may very well happen for efficiency-related reasons; but it is important to note that the first-order objective need not itself be based on efficiency considerations.9

It follows that there are also reasons for SOEs to co-exist with private market actors in a ‘real’ market. The state might want to, for example, make some good freely available at a loss, yet allow private actors to produce their own version of the good for profit to support public duties it deems essential regardless of market provision,10 or it might want to stimulate private R&D while simultaneously developing “certain capabilities, technologies, and knowledges […] without being limited by commercial considerations” (Willemyns 2016).11 The point is that SOEs are, because of the absence of the efficiency rationale guiding private actors with its invisible hand,12 not so affected by the alluring pressures of the market efficiency mandate. They can forgo ‘winning’ specifically so that other objectives are met.

If anything, SOEs are influenced most by being under a constant pressure to project liberal neutrality, which must entail some care duties as

---

9 Of course, it can be. That is, if a citizenry prefers something, the state may have efficiency-related reasons to pursue the fulfilment of that public preference. It should be noted, however, that this line of argumentation just passes the buck to a question economists are generally unwilling to answer, namely: ‘Where do preferences come from?’

10 Although this is in danger of leaving, as Heath rightly notes, “relatively powerless citizens consuming the public version of the good, people whose complaints can more easily be ignored” (2023a, 52).


12 See for a modern, sophisticated view on this point: Hussain (2023).
well as a minimal commitment to citizens’ equality. According to Ronald Dworkin (1978), a state failing to project equal concern and respect cannot be a liberal one. Indeed, some minimal conception of equality is implied by the very concept of neutrality. Of course, most liberal philosophers have taken this conception to imply the state has at least some measure of distributive responsibility, most prominently among them John Rawls. Famously, his principles of distributive justice are based on the idea that “while justice draws the limits, and the good the point, justice cannot draw the limit too narrowly” (1993, 174). As shown earlier, Heath tends to think this limit is attenuated by the adversarial morality of the market; and, indeed, this is reflected in the market efficiency mandate. However, even if one does not buy this type of high-minded argument, it is abundantly clear that a specifically liberal state requires its populace’s trust to function properly, and that this in turn demands at least a veneer of neutrality in distributive matters (Warren 1999; Lenard 2015).

By dint of the above, it is my contention that even if one agrees with Heath that “state involvement in the economy should be guided primarily by the norm of efficiency” (2023a, 60), it is clear that the state’s constitutional commitment to liberal neutrality profoundly affects the manner in which it is able to go about meeting this norm. Indeed, states are not firms. More specifically, SOEs that operate within a market environment experience ‘outside boundaries’—or moral limits—beyond just the task of maximizing profits in ways that allow Pareto efficiency to theoretically obtain under conditions of perfect competition. Indeed, the “extremely onerous constraints that are imposed on [state] procurement” (46) to which Heath appeals, far beyond what can be expected of any private firm, are clear proof of this phenomenon. It should be obvious that SOEs are often prohibited from “developing strong relationships with suppliers” (46) not because of efficiency-considerations, but because of neutrality-considerations. Indeed, SOEs must operate not as if they were operating in a market under conditions of perfect competition, but as if they were operating in a schmarket. And as Heath rightly ascertains, this constraint often puts the state at a significant disadvantage compared to the private sector vis-à-vis profitability. On efficiency grounds, this disadvantage can

---

13 This pressure is much attenuated in SOEs that are only partially owned by the state. I do not go into such cases in this paper.
14 The particulars of liberal neutrality vis-à-vis distributive considerations are significantly more complicated; see, for instance, Schaller (2004). I do not take a position regarding any of this in this paper, however.
DILLINGH / REPOLITICIZING PRIVATIZATION

certainly be said to furnish a *cooperative cost*; yet we can now start to consider the possible *competitive social costs* that may follow suit.

To see this, consider again that firms may justifiably pursue profit due to the competitive social norms that the market appeals to; indeed, on Heath’s account, “markets privilege efficiency over equality in a way that is far more radical than what one can find in any other social institution” (Heath 2023b, 37). However, as we have seen, under conditions of imperfect competition—which is indeed almost always—market competition will only bring about efficiency of the Kaldor-Hicks variant. Importantly, SOEs’ schmarket-constraints *do not allow for such shortcuts*; indeed, whatever the objective being pursued, the pressures of liberal neutrality demand SOEs do not disadvantage actors in the way that Kaldor-Hicks efficient outcomes imply. They are principally not allowed to privilege efficiency over equality to the same extent that ‘pure’ market institutions do. In general terms, we may posit that SOEs cannot as a rule set back one party and privilege another, whilst remaining agnostic on whether the losing party will ever be reimbursed. This would, in a very direct sense, come into conflict with the minor (distributive) equality considerations associated with liberal neutrality.

All this is a rather technical way to spell out that where market competition between private actors experience competitive pressures to produce Pareto efficiency only *indirectly*, state-led economic endeavors experience cooperative pressures to pursue it *directly*. Call this the *schmarket efficiency mandate*.15

**IV. THE COSTS OF USING THE PRICE MECHANISM**

So what does the difference between the market efficiency mandate and the schmarket efficiency mandate entail for the concept of competitive social costs? In this section, I show that the difference between these mandates can be expressed as a competitive social cost, and that this cost is, subsequently, best expressed in terms of *inequality*.

Due to the market efficiency mandate, efficiency gains are, again in general terms, *ex-ante* more *unequal* when the private sector is involved. This is actually a rather uncontroversial point. Indeed, Heath himself notes that “courts often use a KH [Kaldor-Hicks] conception of efficiency in considering competition law cases. Similarly, regulation is often

15 Of course, I certainly do not intend to deny particular problems that may occur in, or are associated with, specifically state provision, such as those that are flagged in Heath (2023a).
justified through cost-benefit analysis (CBA), and CBA uses the KH standard” (Heath 2019, 23). Thus, very often, only *potential* Pareto gains are achieved by the market.16

My additional, more novel contention is that SOEs, conversely and due to the pressures of the schmarket efficiency mandate, tend to more often achieve efficiency gains that are ex-ante more equal. Far from trying to deify SOEs, I must immediately add that these gains are frequently not even *potential* Pareto gains. That is, they are, owing to several difficulties Heath spells out with acuity in his paper, generally much lower than the Kaldor-Hicks gains made by the private sector. Yet due to the schmarket efficiency mandate, they are more often made with an internalized distributive factor that bars SOEs from facilitating the win-lose interactions that characterize the Kaldor-Hicks standard.

The argument, here, is that this comparison reveals a competitive social cost involved in decisions of privatization. Moreover, since both the public and private sector are realistically unable to achieve perfect Pareto optima, but only the private sector is allowed to generate win-lose interactions, this cost is best expressed as one of *ex-ante* inequality. In somewhat more morally laden terms, the Pareto standard clearly internalizes some notion of basic equality; that is, winners may only win if it is not at the expense of other participants. Conversely, Kaldor-Hicks gains constitute only potential Pareto gains; that is, winners may win at the expense of others as long as they are able pay them back with their winnings. Yet it is this very potentiality that constitutes the notion of basic equality furnishing the *normative* facet of the Pareto principle. Indeed, without this facet—if the losers are not reimbursed—we are left with an economic interaction that just generates inequality.

Note that this comparison does not, in fact, incur the levelling-down objection. The oft-used objection attacks the monistic egalitarian notion that only equality is intrinsically valuable. It follows, so the objection suggests, that to egalitarians a more widget-equal distribution is preferable to a less widget-equal distribution in which all actors nonetheless possess more widgets.

---

Egalitarians, then, should prefer B over A. Hence, egalitarians are committed to *levelling down* gains whenever this would create a more equal distribution. This is, of course, a very hard bullet to bite for egalitarians.

Whatever the merits of this device, it is not applicable to my comparison. That is, I am comparing instances that follow the schmarket efficiency mandate in which gains are low, but no actor is ex-ante set back, against instances that follow the market efficiency mandate in which gains are high but, crucially, at least one party is set back.

Thus, we might appropriately express the competitive social cost of privatization as *ex-ante* inequality. Yet the operative word here is, of course, ‘ex-ante’. Heath might even agree with the above analysis wholeheartedly, but go on to argue that the *ex-post* gains—that is, actual gains after welfare state Pareto-enhancement—are such that even a non-ideal perspective on the market justifies the Kaldor-Hicks efficiency resulting from the market efficiency mandate. In fact, Heath might say, the ex-ante competitive social cost, while present, is entirely ameliorated by the Pareto-enhancing effect of a strong welfare state. This is the point, however, where his Market Failures Approach to the state is forced to repoliticize.

---

17 For a most substantive analysis, see section III of O’Neill (2008).
18 Note that the ex-ante prefix is analytically accurate, but may be deceptive. More specifically: on Heath’s account, the Pareto-enhancing effect of the welfare state does not comprise redistributive efforts, but entails pre-distributive and procedural efforts comprising regulation, monitoring, and an encompassing mandate to weed out market failures, generally guaranteeing as close to perfect competition as possible. While there is certainly no t=1 in which such a pre-enhanced outcome is extant, it can nonetheless be
V. Repoliticizing Redistribution

Let us briefly take stock. I have argued that there exist not only transaction costs associated with using the price mechanism, but also social costs associated with affirming the competitive pressures of the market. Moreover, these social costs can be made explicit without reference to deontological considerations and without committing to a socialist political morality; in staying with Heath’s approach, they may simply be expressed in terms of efficiency. More specifically, state-owned enterprises are heavily constrained in the manner in which they may justifiably pursue efficiency gains. They may do so only without setting actors back. The private sector experiences no such constraints, however. Indeed, its very success is predicated on its ability to compete with a lack of such ‘high-mindedness’. The effect of this discrepancy makes it so that SOEs are more likely to produce gains that are ex-ante lower but more equal, whereas the private sector is more likely to produce gains that are ex-ante higher but more unequal.

Now, Heath might think this fine and fair enough, but counter that these ex-ante inequalities are generally mitigated ex-post, via the state’s Pareto-enhancing abilities. This perspective would be consistent with his earlier work on the subject (Heath 2011; 2020). The problem is, of course, that the state’s Pareto-enhancing abilities are directly affected by privatization. Indeed, the main problem with Heath’s account is that it surreptitiously assumes economics and politics are two entirely separable domains, connected only in a somewhat detached, indirect manner. This view invokes the classic, yet now antiquated view of a welfare state run by stern, incorruptible public servants, wholly unaffected by a market that should simply be left alone to work its magic (Marshall 1981; Goodin 1988). Antiquated as it may be (and resist it as he does in Heath 2020), Heath seems now to subscribe to this entirely procedural view, nonetheless.

The state’s Pareto-enhancing abilities are, however, not procedural concepts. Nor are they ephemeral. They are entirely constituted by the state’s ability to pre-distributively monitor, adjudicate and intervene in the economic realm. Here it becomes imperative that we add to Heath’s observation that “state spending as a fraction of GDP typically rose from less than 2% at the beginning of the 20th century to roughly 40% at the cogently expressed as a theoretical object in which the market efficiency mandate has not yet been tempered by the state’s Pareto-ameliorative efforts.
end” (2023a, 27), that an unknown, yet presumably significant portion of that has gone to the private sector via outsourcing and public-private partnerships since the 80s (Lazonick and Shin 2020). This shift has empowered the private sector in a structural sense. More specifically, the shift has made the public sector more and more reliant on the private sector for its very functioning (Farrell 2018). Pressingly, this reliance includes the state’s ability to regulate, monitor and adjudicate the private sector—all that constitutes the state’s Pareto-enhancing qualities.

Concurrently, there is a growing body of both empirical and theoretical work that suggests corporate interests have gotten progressively more leeway in polluting the very political process that is supposed to decide on matters of privatization (Crouch 2011; Acemoglu and Robinson 2019). Synergetic research, moreover, suggests that corporate influence has grown in tandem and requires an approach to market power beyond what the standard economic fare can offer (Claassen and Bennett 2022). This all amounts to the conclusion that “many critical areas of regulation have become characterized [...] by the outsourcing of the development of these rules to the market actors who will be affected by them” (Mazzucato and Collington 2023, 165). And where Heath may counter by saying that his approach would consider these paradigmatically ‘core’ state functions and that they are, therefore, not to be privatized, the real problem only reveals itself when seen through a structural lens. That is, there is every reason to believe that privatizing a previously public institution can realistically severely hamper the state’s Pareto-enhancing abilities in general, even while “the state retains the ability to exercise considerable control” (Heath 2023a, 39) over particular cases. This is so even if a single instance of privatization is deemed to be theoretically ex-post more Pareto efficient, especially when these instances are progressively more often enacted with appeal to corporate—and not public—interests. Indeed, a

skilled government workforce needs to be able to analyse changes both to ensure corporate laws are abided by and to amend or improve those laws where necessary [...] But the more that responsibility for delivering core functions is privatized, the harder it is for the government to ensure in-house regulatory skills are up to date and improving. (Mazzucato and Collington 2023, 165)

The knowledge required for these ‘core functions’, however, is dispersed over various state institutions, many of which are not directly in the business of regulating markets. Indeed, “a loss of knowledge can [...] undermine the state’s capacity to govern relationships with the private sector,
whether through regulation, procurement or other forms of partnership” (164).

None of this is meant to say that these concerns are not relevant without my rigid adherence to Heath’s methodology; indeed, the aforementioned phenomena are explored in much greater detail than I have space for here. Rather, my cursory observations are meant to show that an honest embrace of Heath’s view on privatization taken far enough requires politicization by engaging with these matters. If we, using Heath’s methodology, agree that the difference between market and schmarket efficiency mandates results in ex-ante inequality, we must naturally also analyze what are the processes that affect the state’s Pareto-enhancing abilities on the ex-post outcomes. And, indeed, the analysis of these abilities can only be addressed through structural, political means, beyond a case-by-case approach that appeals to efficiency alone.

Nor is any of this meant to disparage Heath’s meticulous paper. Rather, it is to show that a monist efficiency-based approach to the state is simply unable to be sufficiently exhaustive to cover the phenomenon of privatization. To approach the question of privatization in a depoliticized manner, with reference to policymakers that are “neither stupid nor particularly ideological” (Heath 2023a, 31) is to miss a crucial political effect of privatization that has significant consequences for how that analysis should be conducted in the first place. Indeed, the main theoretical upshot of this paper should be that even a nigh-manic focus on the Pareto principle will eventually be unable to sidestep questions of inequality, because it is a moral principle that internalizes egalitarian considerations. Thus, either Heath’s approach bites the bullet and collapses into a form of aggregate utilitarianism (Moriarty 2020), or it embraces these egalitarian considerations, thereby repoliticizing privatization.

VI. CONCLUSION
I have argued that Heath’s Market Failures Approach to privatization is made significantly more problematic by the presence of ex-ante inequality. Heath’s analysis, while extremely informative, tends to compare idealized markets with non-ideal state institutions; readjusting this imbalance subsequently suggests that his approach is somewhat morally emaciated. That is, it tends to gloss over exactly why political and egalitarian considerations—pertaining to relations between economic actors—are interesting to questions of privatization in the first place. There is every reason to believe that this view is “commonly held by economists, and
much less commonly held by philosophers and political theorists” (Heath 2023a, 60) because economists are generally taught (wrongly and ruinously) that economics and politics are supposed to be separated analytically. Subsequently, the underlying politics are, unfortunately, left to philosophers and political theorists.

REFERENCES


Savriël Dillingh is a PhD-candidate at the Erasmus Institute for Philosophy and Economics (EIPE), Erasmus University Rotterdam. He works in political philosophy and business ethics, with a particular focus on the ways in which both cooperative and competitive norms influence market processes and institutions.
Contact email: <dillingh@esphil.eur.nl>
Elements for a Normative Theory of Privatization

RUTGER CLAASSEN
Utrecht University

Abstract: Heath’s paper on privatization defends a broadly welfarist-economic approach in thinking about the legitimacy of privatizations. This approach is ‘instrumentalist’ (in contrast to deontological approaches). In this response, I accept the value of an instrumentalist approach to privatization, but argue against Heath’s welfarist version of it, and argue in favor an alternative. First, the ends we seek when thinking about socially vital goods (our theory of public interests) should go beyond Pareto-efficiency. Second, as to the means we employ to realize these ends, we need a view of markets which takes into account not just their competitiveness, but also the distribution of power. This means we need to differentiate market types. Third, we need to differentiate ownership types beyond the standard shareholder-owned company. Alternative ownership structures may be able to realize public interests more easily and hence make privatization less problematic. On these three issues, the picture Heath sketches leaves out too many of the variables that, in the end, may be decisive in whether or not a privatization is acceptable.

Keywords: privatization, public sector, markets, ownership, Joseph Heath

JEL Classification: D4, G32, G38, H4 J54, L2, L3, P16

I. INTRODUCTION

Joseph Heath’s paper on ‘anodyne privatization’ is—as so much of his writing—a provocative and enjoyable read. Slinging his arrows at fellow philosophers and political theorists, he argues that the bulk of the field has been—similar to public opinion in many countries—in the grip of an overly pessimistic view on privatization. The basic attitude is one where whatever the state does is salutary, while whatever the market does is a sell-out to evil capitalist forces. Against these biases, he urges readers to
look at the question of privatization in a less ideologically charged way. Some privatizations are ‘anodyne’, and should be accepted as such. With the economics-inspired welfarist theory he has defended at length elsewhere (Heath 2014; 2020) in the background, he discusses how there are often good economic arguments to accept privatization.

While reading the paper, I found myself agreeing with much of it. Part of the reasons for that is that Heath makes many acute observations, which do help the debate forward. For example, I fully agree the philosophical debate has often (though not always) focused on public services that belong to the “core coercive apparatus” of the state (Heath 2023a, 27–28), while what we (also) need is a look at the more standard economic goods and services often provided by the public sector. It is good that Heath takes on this task in his paper and hence invites other philosophers to do so too. I also found the analytical device of distinguishing different types of privation incredibly illuminating (see the figure at Heath 2023a, 37). As Heath convincingly shows, with helpful discussion of examples, these different types of privatization force upon us different sets of considerations to ponder. Finally, I agreed with the overall conclusion. Heath’s main message is that “privatization of certain state services, in certain cases” (26, emphasis added) is anodyne. In my view, it is hard to argue (unless one is a state socialist) against such an incredibly modest conclusion. So what does this leave us with?

If we accept the main message, we accept privatizations range from extremely problematic on the one end, to completely anodyne on the other end. We then can move on to the really interesting question: how to think about the large field of cases in the middle, where things are less clear-cut? For these ‘hard cases’, and even for identifying the extremes, we need a theory. As Heath himself recognizes at the end, we need “a conceptual framework, or grid of sorts” (59) to help us think about all the cases where things are less clear. Now it is somewhat unclear to me whether Heath takes himself to have developed such a theory in his paper. But I take it that his mobilization of an welfarist-economic approach throughout the paper (36, 60) implies that he would argue that this is the best candidate for the job. In my response I want to discuss whether this is right. Which theory for deciding privatization do we need?

I do not come to these questions with a blank sheet. Elsewhere I have argued in favor of (at least the rudimentary outline of) such an approach (Claassen 2017). Here I want to use the occasion to update and expand on my earlier approach, in confrontation with Heath’s stimulating paper. I
will start by giving a more detailed sketch of the philosophical landscape of positions on privatization (section II). After that I will confront my approach with Heath’s on three issues: the identification of public interests (section III), the design of markets (section IV) and the role of ownership (section V). On each of these issues, I will suggest that Heath’s welfarist approach does not give us what we need, and I will argue for a contrasting perspective. Now I do not think this will decide the debate, since Heath may argue that many of the contrasting considerations can be absorbed into a welfarist theory (the absorptive capacities of welfarist-economic theories are renowned). But such acts of absorption will render the welfarist theory into more of a complex hybrid, and then we need to know how the relations between the different parts within such a hybrid would be structured. Whatever of the prospects of welfarism, I hope my remarks will help the debate further, towards the development of normative theories of privatization (welfarist or not) sufficiently rich to tackle the relevant ‘hard cases’ on their merits.

II. THE PHILOSOPHICAL DEBATE ABOUT PRIVATIZATION

The field of philosophical approaches to privatization is usually divided into two camps: instrumentalist versus non-instrumentalist (Cordelli 2020, 31–42; Schwartzberg 2018, 1), outcome-based versus agency-based (Dorfman and Harel 2021, 1–2), or deontological versus consequentialist approaches (Heath 2023a, 26–31). On the one hand, privatization can be evaluated by judging it based on its outcomes, comparing outcomes under privatization with outcomes under a regime of public provision. A wide variety of criteria, economic (welfare, efficiency) and justice-based ones (egalitarian norms of distribution) can be used as outcomes. These theories lead to contingent statements on privatization’s rightness or wrongfulness, since they depend on empirical generalizations about the effects of privatization. On the other hand, privatization can be evaluated on the basis of a priori arguments about what should or should not be the proper realm of private and public activity. These theories deliver necessary classifications, independent from empirical contingencies. Of course, this dichotomy is a simplification, but I will accept it here as a sufficiently accurate sketch of the field.¹

¹ For example, Dorfman and Harel (2021) distinguish a ‘process-based’ approach as a third category, which I ignore here. And the application of the distinction to specific authors is always contested. For example, Cordelli (2020) classifies her own approach as transcending the instrumental/non-instrumental divide, while Heath classifies her theory as deontological (see also next note).
In his contribution, Heath argues that whatever the merits of deontological theories for certain core state functions (like law-making and judiciary functions, the military, and prisons), they are unsuitable for evaluating the economic functions of the state, such as the delivery of public utilities and welfare state services. As he states: “there are no global arguments for or against privatization” (41). He emphasizes how many lower-level services in the public sector are regularly and uncontroversially contracted out. As an antidote to deontological theories, Heath’s message is that there are many unproblematic cases of privatization. Decisions need to be made pragmatically, by judging each case on its merits. Like Heath, I believe an instrumentalist approach is the way to go in thinking about the economic functions of the state. But I will give this approach a different shape.

Instrumentalist approaches have two key features. They are good-specific: they do not target the aggregate level of privatizations throughout the whole public sector, but aim to evaluate privatizations on a case-by-case basis. Considerations about the particular nature of, say, higher education, water, energy, or social work, play an important role in instrumentalist approaches. Second, instrumentalist approaches are comparative. The privatization question asks us to make a comparison between public and market provisioning when a transition from the former to the latter is contemplated. But in principle, the comparative question also applies when a reversal (from market to public provision) would be considered; and it also should include non-market, non-state forms of provisioning, such as the ‘commons’ which some argue are a third alternative. The privatization question is but a special case of the more generic question of economic organization: by what mode of provisioning best to organize the production of good x?

The structuring of the field into deontological approaches on the one hand and instrumentalist approaches on the other hand, bears a strong resemblance to public debates about privatization. At least from my experience of following these debates in my country (the Netherlands), we find, on the one hand, economists and economically-inspired policy makers arguing in favor of a specific privatization. Economic theory offers well-established theories of market failure and government failure (public choice theory), and hence has great intellectual authority in these debates.

2 Hence I will say nothing about the philosophical merits or problems of the deontological approaches that are his primary target. For my view of Cordelli’s theory, see Claassen (2022). There I also discuss whether Cordelli’s view is indeed ‘global’, as Heath suggests, or if it relies on empirical considerations.
On the other hand, those critical of privatization often put forward their objections based on intuitions that privatization in this-or-that case ‘just would be wrong’, without a backing into something closely resembling a systematic theory of the kind economics offers; or alternatively, they also use economic theory to try to show privatization would be inefficient. Overall, this makes the impression there just is ‘no other game in town’ than economic theory.

To me, both sides are unsatisfactory. What we need is the project of an instrumentalist theory which allows for a systematic set of considerations, some of them independent from economic theory, others including but going beyond economic theory. Such a normative theory would have the potential of offering a rival to economic theory narrowly conceived. One of its additional virtues would be its ability to reintegrate some of the motivating concerns of deontological approaches into its instrumentalist framework. For such a project, there are various directions. In my (2017) paper I conceived of it along the following lines. To think about privatization of a specific good/service, we need to distinguish ends and means. The ends of any system of provisioning for good x are the interests with respect to a good x that should be realized. A normative theory of public interests (or values, principles, rights), whether or not made explicit, is the source of one’s judgments about this. The means are a range of possible institutional arrangements (private or public). They must be evaluated in light of their (expected) merits to fulfill the ends. In my framework, we need to take five steps, the first two steps being about the ends, the next two about the means. The final step concludes.

Step 1: a normative theory of goods. For any good, we need to decide whether we think the state—as a matter of justice—should be concerned with its provision at all. I propose to think of this in terms of social rights: to which goods do citizens have a social right? We may come to a positive answer for some (for example, health or education), a negative answer for others (to wine or yachts). Only with a positive answer, do we need to move to steps 2–5. Various theories of justice

---

3 To clarify: I am not suggesting the deontological approaches in the philosophical debate about privatization are analogous to these intuition-driven arguments. The similarity is in the categorical stance they take in rejecting privatization, not in the sophistication of the arguments used.

4 I will not analyze the overlaps here, but roughly, I think the concerns about legitimate authority and the usurpation of public power that recur in deontological theories link up with the second main point I make in section IV, about the mutual influences between politics and economics.
could do the job for this first step. For example, a Rawlsian theory will cash this out in terms of a list of primary goods; a basic needs theory in terms of a list of basic needs; a capability theory in terms of a list of central capabilities; a human rights theory in terms of a list of human rights.\footnote{For purposes of my argument, I will be agnostic between many such theories to make this paper’s argument broadly acceptable. I suspect that many of these theories converge in their support for a basic normative requirement, for a standard list of public goods. For my own favored normative theory, see Claassen (2018).} The theory-application will issue in a judgment of the form: ‘in a just society, all citizens have an entitlement to enjoy good \( x \)' (where \( x \) is a public good or service, like health care, education, housing, water etc.). I will call this the \textit{basic normative requirement}.

Step 2: a normative theory of reasons for considering the value of private versus state provision. The basic normative requirement implies that a state needs to take adequate care that all citizens can enjoy \( x \). But it does not always have to provide these goods themselves. To think about this, we need a further set of criteria. Here I proposed a liberal framework, arguing to think about three separate criteria:

(2a) freedom/autonomy. As a default citizens value freedom of choice; that is to say, being able to choose between providers. But for certain goods citizens may lack the ability to make good choices.

(2b) inequality. Inequalities in provision are predictably larger under conditions of market competition than under state provision. One may regulate to mitigate this, but this typically undercuts the dynamic effects of markets as well.

(2c) efficiency versus inefficiency of market versus state provision. Markets have efficiency benefits when certain conditions are met. But whether they can be met is contingent (see also step 3).

These two steps offer an a priori set of reasons, which need to be applied to a specific good, in light of our more specific normative commitments. For example, depending on one’s commitments, one person will judge inequalities in health care between citizens to be more acceptable than another person. But the final choice will not just depend on the weightings of these various factors in the realm of pure normative theory. For we cannot presuppose that the privatization of, say, a certain health care service, will lead to ‘result \( x \)' on the freedom dimension, ‘result \( y \)' on the
equality dimension, and ‘result z’ on the efficiency dimension. This also depends on the markets that we get (as a result of governmental design and enforcement on the one hand, spontaneous processes on other hand). Hence we need to consider:

Step 3: differentiations between markets: not all markets are alike (more on this in section IV). How we expect markets to perform on the three reasons identified in the previous step, depends on what markets we will get when privatizing them. This in turn depends on what is achievable politically.

Step 4: democratic decision-making in service provision. For many goods/services, even when they are privatized, there is reason not to ‘simply privatize’, but to embed the privatized service in a context in which service users and sometimes also citizens at large can raise voice, not just ‘exit’. Designing appropriate mechanisms of voice is the institutional challenge at this step.

Steps 3 and 4 together provide a picture about what a privatized system would offer, at its best. I presume a similar exercise is/has been made about the existing state-based system. Together, this shows:

Step 5: application to come to a decision: a democratically legitimated body needs to make a decision, by combining the criteria under step 2 with the market designs options under step 3 and 4 (including an estimation of how realistic these are) as well as knowledge of the existing public system. It ends up in favor or against the privatization of a particular good, under a particular institutional design.

In light of this framework, I will now delve more deeply into the question of how to determine the public interest (section III), what theory of markets to adopt (section IV), and the question of ownership (section V). This will provide me with the occasion to refine and modify steps 2, 3 and 4, respectively, and confront Heath’s theory at every stage.

III. HOW TO DETERMINE PUBLIC INTERESTS WITH RESPECT TO A PUBLIC GOOD/SERVICE?
How to develop a satisfactory normative theory of the public interests in question? Economic theories rely on efficiency as the guiding norm, an
approach adopted by Heath throughout his work (Heath 2011; 2014). The operationalization of this norm happens through the theory of market failure. The perfectly competitive market is the theoretical benchmark; there is only a rationale for public provision if markets fail to work efficiently.

Efficiency is a tricky concept. In the abstract, it is a formal norm comparing arrangements which aim to satisfy multiple ends: “an allocation of resources is efficient if it is impossible to move toward the attainment of one social objective without moving away from the attainment of another social objective” (Le Grand 1990, 559). As Le Grand argues, this is incompatible with seeing efficiency as one of these primary objectives itself. Efficiency is a relation between two (or more) primary objectives. None of these objectives is efficiency itself. Under this formal understanding, the infamous equity/efficiency trade-off is incoherent, since it treats efficiency as a norm and opposes it to equity (Le Grand 1990, 560). The trade-off should be understood as referring to a substantive interpretation of efficiency, such as Pareto-optimality. A situation is Pareto-optimal if the well-being of one person cannot be improved without reducing the well-being of others. Thus understood, Pareto-optimality includes the acceptance of a particular norm of equity. What gets traded off is the equity embodied in a Pareto-optimal arrangement against equity as understood by any other norm of equity, informed by one’s favorite theory of justice (Le Grand 1990, 566). Pareto-optimality is just one example of a substantive efficiency norm, but it gains particular salience given its popularity in economics.

In moving forward, then, we need to decide whether to accept Pareto-optimality as the only relevant norm to identify public interests or as a norm that needs to be complemented with and where in conflict, traded-off against, non-welfarist norms of social justice. Heath is a fervent defender of exclusively Pareto-based theorizing in other contexts, such as the debate in business ethics about the moral responsibility of firms (Heath 2014). By contrast, Singer (2018a) and Blunden (2022) have both argued that Heath’s market failure framework needs to complemented with non-welfarist norms. What about the privatization question? Here too, Heath explicitly refers to his own framework as based on market failure theory (Heath 2023a, 36, 60).6

6 The issue is bit more complicated, because Heath does acknowledge at a few places in the text the moral relevance of egalitarian considerations. For example, he is concerned about costs “borne disproportionally by the disadvantaged” (2023a, 52) But he does not offer a systematic view about the relation between justice norms and efficiency-based
Now this is obviously in contrast to those authors who, in the debate about privatization, have proposed justice-based theories (for an overview, see Cordelli 2020, 24–33). These authors often combine justice with efficiency; the result is a Pareto-efficiency-cum-justice theory for determining the public interest.⁷ In practice, such theories support the endorsement of many privatized arrangements (for their efficiency gains), and then supplement market-based provisioning with state subsidies to poorer consumers to prevent unjust outcomes. This is a pragmatic, voucher-based strategy to combine the efficiency virtues of markets (when present) with a concern for mitigating inequality (when it works). Should we choose between Heath’s purely-Paretian approach or such Pareto-plus-justice approaches? In the following, I want to propose yet another way forward, which leaves Pareto behind. It does integrate efficiency and justice considerations, but in a different way, making use of the formal, non-Paretian sense of efficiency mentioned above.

In a variety of contexts, practitioners have come up with what they call a ‘public service trilemma’. The idea is that three often-recurring values in public service provision pull in different directions: affordability, accessibility, and quality.⁸ For any of these three values, picking two of them puts pressure on the third one (see figure 1 for illustration).⁹ Take health care as an example. First, one can design a health care system that is accessible to all, with low eligibility thresholds and no private co-payments, and that delivers high-quality care per patient. Such an extensive,

---

⁷ We can distinguish a weak and a strong form of such combined theories. On the weak version, justice norms are merely used to decide between Pareto-optimal allocations (hence where Pareto itself is indifferent), on the strong version justice norms can push the theory towards a Pareto-non-optimal arrangement. In practice, such theories support the endorsement of privatized arrangements (for their efficiency gains), and then supplement market-based provisioning with state subsidies to poorer consumers to prevent unjust outcomes.

⁸ These three values here provide a revision of step 2, in that they can show us in more detail how to apply the criteria of equality and efficiency (I leave freedom/autonomy out of consideration here).

⁹ Years ago, I came across this trilemma in a policy document. Trying to relocate its origins in the academic literature for inclusion in this paper has been a hassle. In higher education, the original source seems to be (Ansell 2010, 165), who refers to ‘level of enrollment’ (accessibility), ‘degree of subsidization’ (quality) and ‘overall public cost’ (affordability). In health care, independently, the trilemma emerges as ‘the iron triangle of health’ in Kissick (1994).
high-quality system is going to be expensive to the public budget, however. This will crowd out what can be spent on other public goods. Second, if one instead prioritizes a more modest public budget for health care (to diminish this pressure on the public purse), and also wants to maintain its accessibility to all, then one is bound to cut back on quality. Third, if one wants to avoid quality deterioration but hold on to the affordability of the system, one must make the system more selective, by designing higher eligibility criteria and/or co-payments.

The public service trilemma is just a simplified framework. But it forces policy makers (and normative philosophers) to move beyond a very generic normative basis (that x is a basic right, primary good, central capability, etc.), to something more precise: a view of how—from the perspective of one’s favorite normative theory—trade-offs between at least three public values with respect to the provision of x need to be made. None of these three values simply represents ‘equity’ or ‘efficiency’. Instead, ‘equity’ and ‘efficiency’ each get a different place in this alternative framework.

Equity considerations motivate a concern for all three values of quality, affordability, and accessibility. A concern for accessibility is a concern about scope: to spread access to a particular public service to a wider range of citizens. It is almost self-evident that this is an egalitarian
consideration. Perhaps less obviously, the value of *quality* is also egalitarian: it relates to the fact that the basic normative requirement to provide health care, education, etc. is always to provide it at a certain *level*. A poor product does not suffice, an egalitarian wants to get the target population to the quality level normatively required. Finally, egalitarian considerations motivate concern for *affordability*: how much of good x can be provided while at the same time leaving public funds for other normatively required public services (at their own normatively required levels)? Egalitarian justice as a generic norm pulls in various directions when thinking about public services. More specification is needed to come to a particular balancing of various egalitarian concerns in the situation of the public service trilemma.

The role of efficiency is also reconceived, compared to theories that adopt Pareto-optimality as the overarching norm (with or without supplementary justice considerations). In the trilemma framework, efficiency refers to the extent to which these various public interests can be simultaneously satisfied, hence how sharp the trilemma is. Efficiency returns to its formal meaning, of a handmaid, preliminary and subservient to the other three values, which are primary ‘social objectives’, in Le Grand’s terms. It may (or may not) be possible to increase performance of the system on one value (say, accessibility) without paying the price for this in terms of one of the other two values. If so, the system is inefficient. But when no further increases in one value are possible without decreases in (one of) the other two values, the system is efficient (these judgments can change over time; for example, lacking uptake of technological innovations may turn efficient systems into inefficient ones). Efficiency analysis helps the public interest-theorist, by informing them how sharp the trade-offs between the fulfillment of the three public values are expected to be in practice.

Now one could object to this that what is doing the work here is still a norm of Pareto-efficiency, but now at a meta-level, in the consideration of the relation between the three values. In reply, two things. First, this is true, but as an artefact of the fact that I adopted Le Grand’s definition of the formal notion of efficiency (see the quote above), as one in which a system is efficient when it cannot improve on objective A without a setback in objective B. One could equally adopt a definition of the formal notion of efficiency along Kaldor-Hicks lines, so that a system is efficient when the gain in objective A leads to a comparatively smaller set-back in objective B. Second and more importantly, the three values over which
the efficiency calculus is made are themselves motivated by egalitarian—not welfarist—considerations, as noted above. And most importantly, efficiency here does not define the public interest in question, but it merely helps us to see to what extent we can escape a trade-off between these values. It does not help us in making the choice between them, to the extent that such an escape is unavoidable.

Years ago, I saw this trilemma live-in-action, when a room full of academics were passionately debating the budget cuts the government wanted to impose on our (publicly funded) university. Some said: 'well, if this is the money the government is prepared to give to the university, we should accept that students get lower quality' (that is, less attention/instruction from their teachers, given that a smaller budget allows fewer teachers, hence a lower teacher/student ratio). Others said: ‘we should provide the same quality, but to fewer students. Let us make academia more selective and reserve higher education to a smaller part of the population’. Still others said: ‘we want to maintain accessibility and quality; so we should strike, not accept the budget cuts, press government for more funds’. All of these are reasonable positions. Efficiency cannot decide between them. Because the government did not take back the cuts, and did not allow universities to be more selective in their admissions, and because dedicated university teachers often hate lowering quality for their students, the only way forward was to search for ways to (somehow magically) work ... more efficiently.

Heath briefly mentions education when discussing type-1 privatizations, noting that in a public system not charging (sufficiently high) user fees creates congestion (student crowd classes by attending); and that it leads to cross-subsidization between students who derive variable benefits from their (philosophy versus medicine) diploma’s. I agree with these as factual statements. But he presents these phenomena as reasons for privatization. But how so? It would seem to me that in both cases, we can see that a free market for higher education will solve the issues of cross-subsidization and congestion differently from a publicly provided system. Take congestion. Variable fees in a free market will solve congestion by raising the price. Keeping quality constant, some parents/students will drop out because they cannot/do not want to pay the price; others will sacrifice their savings or take out loans. The trilemma is resolved through the market mechanism, which leads to Preto-optimality (if all goes well), as well as the realization of its own concept of equity (as we saw Le Grand
argued).¹⁰ Heath confirms this when he suggests the ‘general principle’ that—under certain conditions—citizens should pay according to their benefit from public services (2023a, 45). But the other option is to solve the problems through a public system, and then the outcome will depend on political convictions about which values should weigh more heavily. When congested, the government may decide to accept lower quality (‘good enough’). Or to increase taxes and increase expenditures, to maintain quality. Or to restrict access. Anyhow, the choice between these market and state-based ways of solving the problems mentioned by Heath, cannot be made by reference to efficiency. His welfarist framework diagnoses the problem of congestion, but—as far as I can see—does not offer reasons for the superiority of the specific route of privatization. That route embodies its own, implicit, trade-off with respect to the trilemma’s three egalitarianism-motivated values. It may, or may not, be superior to the alternatives.

The trilemma framework asks us to make a public judgment about the trade-off between the relevant values; this is what will be ‘the public interest’ in the matter at stake. But I hasten to add that this normative perspective does not imply a bias in favor of public over market provisioning. The normative theory is about the ends, not the means. It is fully compatible with (regulated) markets, if these prove to come closest to realize the value trade-off made in the theory. But this is different from letting market forces make the choice for us. Why insist on this social planner’s perspective about the choice with respect to the trade-off? Ultimately, the reason for this is that the goods in question are a matter of justice (that is, qualifying as generating a ‘basic normative requirement’ of justice, per step 1 of my framework).

This perspective is also, I take it, in line with the dynamic of public debates about these matters. In the last decades, many Western countries public services were privatized but—perhaps unsurprisingly—public concern about these services has not diminished. As soon as a privatized system falls short of meeting public expectations, the government is criticized for failing to regulate the market to meet these interests. Although private providers are legally responsible for failures to live up to public standards, the public continues to address their complaints to parliaments and government officials. Apparently, there is a non-market moral

¹⁰ And, Le Grand adds that it is a “particular, and to many a rather unattractive, concept of equity, one in which a greater value is placed on increases in the utility of the better off than on similar increases for the worse off” (1990, 565).
baseline—which is continuously re-established through public deliberation and contestation—about what ‘we the public’ expect (in terms of affordability, accessibility, and quality) when it comes to specific public goods and services.

My aim in this section was not to argue in favor of a particular balancing of the three values in the public service trilemma. I am not even claiming that these three values are exhaustive (although I think they do cover a lot of what is usually discussed in cases of privatization). The aim of this section was more modest, that is to offer an alternative for theories that use Pareto-optimality as one of, or even the only, relevant norm for determining the complex balance of values we call ‘the public interest’.

IV. Market Design, Power, and Politics

So far, I have made use of a rather undifferentiated picture of ‘the market’. But now we need to move beyond this. To make decisions about a scheme of privatization, we need to know the prospects for markets to protect the (balance of) public interests, whatever we judge it to be. Such an inquiry could be conducted in purely economic terms, along the lines of a theory of market failure; but in this section I want to suggest we need a broader view, which in addition to market failures stresses two themes: 1) markets are always embedded in a set of social and legal norms; each of which leads to different market types where different groups are (dis)empowered; and 2) marketization may affect the ability of politics to set and enforce these norms in the first place, resulting in a vicious circle.

IV.I. Market Design and Power: Differentiating Market Types

To introduce the problem, let us start from what I take to be a key passage in Heath’s paper:

It is certainly the case that highly competitive markets tend to promote firms with efficient internal organizational structures, just as markets with low barriers to entry are very effective at rewarding innovation. But it is not the private ownership structure of firms that is generating these beneficial effects, it is the external environment in which they are operating. Some of the greatest disappointments in the history of privatization have occurred when states sold off assets to the private sector, but failed to create a competitive market for those firms to operate it, leading to the grim discovery that private monopolies can be just as inefficient as public monopolies, and often more
infuriating to deal with. Thus the promise of efficiencies arising merely from the change in ownership structure winds up being a free lunch (that is, an illusory benefit). (2023a, 40)

Two points emerge from this passage. First, the insufficiency of ownership thesis. A shift of ownership is insufficient to bring about the often-promised efficiency improvements of privatization, since these benefits only materialize if the external environment is organized as a competitive market. Second, then, the necessity of competitive markets thesis. Privatization requires competitive markets. In this section, I concentrate on the latter (ownership follows in section V).

Heath relies on standard economic theory, in particular the theory of market failures. Successful privatization requires market design. As Heath suggests, states can fail at this job—which is when we get ‘infuriating’ private monopolies. This hints at theories of government failure, which Heath does not explicitly discuss in his paper. Market and government failure are two necessary parts of any comparative evaluation of the merits of market and state provision. A comparison between an idealized market model and a moderately successful (‘realistic’) public provider (or vice versa) is unfair and unhelpful. The early waves of enthusiasm about privatization in the 1980s and 1990s very much relied on a lack of experience with the difficulties of market design; a few decades later, a more sobering attitude prevails amongst many policy makers. Realism about markets is realism about their failures plus realism about government's abilities to remedy these failures.

My suggestion here is that we need to move beyond competitiveness as the only relevant feature of market design. Competitiveness leads to a differentiation in terms of a spectrum of more or less competitive markets. While this is necessary, it is not sufficient. The relevance of this for the theme of privatization is that there is not one (economic text-book) way to privatize, that is to establish competitive markets. There is a multiplicity of design choices, each of which consists of legal norms, leading to various outcomes on other dimensions than competitiveness, most notably the power of different groups. Markets are always socially and

11 Moreover, Heath rightly points to a second component of the theory of markets needed for assessing privatizations: the theory of efficient contracting. Much of the problems in privatizations arise, as he notes, from problems in contracting between outsourcing states and market parties (2023a, 45–47, 51, 54).
legally embedded, in informal and formal norms which determine the allocation of power between market participants. As Jane Gingrich argues:

Markets in public services will never match a neoclassical model of perfect competition [...]. However, there is no single ‘second-best’ public service market. Several structural features of public services create significant trade-offs among different modes of distribution and competition, and the concomitant incentives that they create. (2011, 8)

To deal with this plurality of options, we need a theory displaying the major design options and their normatively salient features. Gingrich provides a good example. She studied how privatization experiments in various countries can and have empowered different groups, depending on the power constellation in politics when privatized. She distinguishes two dimensions: production and allocation. The allocation dimension is about whether public regulation and financing of markets lead to selective access to the neediest, or more robust universal access to a larger group of users. The production dimension is about the structure of competition, and the extent to which it empowers the three groups involved (the state, users, and producers) to realize their preferences about the balance between costs, quality, and profits (for summary, see table 1 below). States

<table>
<thead>
<tr>
<th>Allocation Dimension: Responsibility for Access</th>
<th>Production Dimension: who has Effective Control?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collective</td>
<td>State: efficiency aims</td>
</tr>
<tr>
<td></td>
<td>Managed Market</td>
</tr>
<tr>
<td></td>
<td>Users: Quality aims</td>
</tr>
<tr>
<td></td>
<td>Consumer-Controlled Market</td>
</tr>
<tr>
<td></td>
<td>Producers: Profits and rents</td>
</tr>
<tr>
<td></td>
<td>Pork Barrel Market</td>
</tr>
<tr>
<td>Recent English contracting in education</td>
<td>Swedish health care market in the early 1990s</td>
</tr>
<tr>
<td></td>
<td>English elderly care market in the 1980s</td>
</tr>
<tr>
<td>Individual</td>
<td>State: efficiency aims</td>
</tr>
<tr>
<td></td>
<td>Austerity Market</td>
</tr>
<tr>
<td></td>
<td>Users: Quality aims</td>
</tr>
<tr>
<td></td>
<td>Two-Tiered Market</td>
</tr>
<tr>
<td></td>
<td>Producers: Profits and rents</td>
</tr>
<tr>
<td></td>
<td>Private Power Market</td>
</tr>
<tr>
<td>Dutch health care markets</td>
<td>English education market</td>
</tr>
<tr>
<td></td>
<td>English elderly care market since mid-1990s</td>
</tr>
</tbody>
</table>

Table 1: Variation in market types. Reproduced from Gingrich (2011, 12)

---

12 The social embeddedness of markets is a longstanding theme in economic sociology, following the groundbreaking work of Karl Polanyi (2001) and Mark Granovetter (1985). More recently, legal institutionalists have taken up the theme of legal embeddedness (Deakin et al. 2017; Pistor 2020), following the lead of older generations of economic institutionalists (Commons 1995).
want low-cost services, users want high-quality services, and producers want profits. The result is six distinct market types, depending on the design choices made (see table 1). Perhaps the details of Gingrich’s classification scheme can be improved upon; I am not wedded to any of the particulars. Here the scheme serves as an illustration of how a power-based analysis leads us to consider multiple market types, which each have different outcomes. The market is not a neutral conveyer of pre-existing inequalities in power and wealth, but can enlarge or diminish them, depending on the design (Dietsch 2010).

This attention to power-differentiated types of markets is largely absent from Heath’s discussion. A good illustration is his treatment of type-II privatizations, where the state contracts out (2023a, 45–51). Here too, he launches a general principle, that is to say “if the market is reasonably efficient, the egalitarian objectives can be achieved through monetary transfers” (50). This motivates a reliance on markets for essentials such as food and housing, where we do care about who gets how much. The discussion motivating this principle relies on examples of the (often overlooked) power-asymmetries within the public sector, where insiders (powerful public officials) can skew public provisioning away from the general interest (50–51; “accumulation of significant resources and economic power within the state”). In Heath’s comparison, however, I found it strange that no attention is paid to similar power-asymmetries that can arise within the market sector, and what that means for our judgments about privatization. This is exactly the kind of analysis Gingrich’s typology puts forward. In Heath’s discussion, the market is presented as attractive based on efficiency (with only an efficiency-based qualifier: ‘if the market is reasonably efficient’), the public sector based on power-differentials.

Again, this focus on public sector power-problems is perhaps because the aim of Heath’ text is to undermine a pro-state bias in his audience. But when we move forward to theorize privatization, both efficiency and power should play a role on either side.

**IV.II. Market Design and Politics: Addressing Political Dynamics**

Market design presumes a social planning perspective. A lot of recent literature, however, emphasizes how market design is an outcome of political processes that are themselves under the influence of market actors. This we need to take into account.
From the point of view of a normative theory (whether created by economists, philosophers, or an expert bureaucratic elite), a certain market design may be optimal. But in reality, market actors have often successfully influenced the process to skew the design to their benefits. A representative example is sociologist Colin Crouch, who documents how and why privatizations have often failed to deliver on their promises. The political influence of business is key to the explanation (2011, 71–96). There is by now a large literature by economists explaining tendencies to concentration in markets and their political effects (Davis 2022; Eeckhout 2021; Philippon 2019). Theories of rentier capitalism describe how corporations with market power extract rents from others in the economy. Protection of their ability to control key assets which deliver these rents is a key part of the explanation; and part of this ability is exercised through political influence (Piketty 2020; Christophers 2020; Sayer 2015). Perhaps the most impressive achievements in understanding the feedback loops between political and economic actors, come from economic historians and economists with an interest in long-term history, such as Acemoglu and Robinson (2012; 2019), Bavel (2016), and North, Wallis, and Weingast (2009). In their work, they show how throughout history attempts at rent-seeking by elites have influenced the design of political and economic institutions and their mutual relations: ‘extractive’ institutions when elites win; ‘inclusive’ institutions when a broader set of the population successfully mobilizes against this.

Against this background, I cannot share Heath’s remarks in his conclusion, that “a great deal of the political rhetoric surrounding privatization is a hold-over from an earlier era, in which social-democratic parties were still committed to the gradual nationalization of the entire economy” (2023a, 59). This makes it sound as if citizens and others (a few sentences later, he adds egalitarian philosophers to the party) who are concerned about privatization are relics from the post-war decades. In my observation, most critics of privatization are motivated by the extraordinary failures of today’s private companies to work in the public interest. This is not an antiquated worry from people growing up during the heady days of flower-power and Che Guevara. Key example are the banks, which were nationalized in some countries in the wake of the 2008 financial crisis since there was no other way to save the financial system from collapse; big tech companies, which drive many politicians and policy makers to madness, in their abilities to invade our privacy, render us addicted to social media, and spread fake information manipulating democratic
processes; and energy companies, a renewed target of pleas for nationalization, since critics see no other way to put them on a track to seriously investing in sustainable energy resources and stop them from putting money into new fossil investments. Some of these companies (or their precursors) were formerly in public hands (such as banks and energy companies). Others would probably have been in public hands, had they existed 50 years ago (such as tech companies).

My point in mentioning these examples is not to suggest that (re-)nationalization of these industries would always be our best option. Yet another solution may be a different, non-extractive type of ownership design (see next section). The point here is that concerns about privatization derive from legitimate worries about concentrations of power and wealth today, and their political effects. To understand how such extractive market designs can (if at all) be prevented, a philosophical theory of privatization would need to incorporate the politico-economic models mentioned above (Claassen and Herzog 2021). As a result, it may endorse conclusions which diverge from textbook economics.

An implication of this is that we may need more controls on producers, to prevent unwelcome political dynamics from market power, than a purely economic analysis would suggest. Acemoglu and Robinson in their work on state-market dynamics make the following observation. In their view, ‘textbook economics’ prefers not mingling with the price mechanism, and redressing any inequalities post hoc through fiscal redistribution. Against this they argue:

But this way of thinking incorrectly separates economics from politics. For the Leviathan to take market prices and the distribution of income as given, and just rely on fiscal redistribution to achieve its objectives might translate into very high levels of taxes and redistribution. Wouldn’t it be better, especially from the viewpoint of controlling the Leviathan, if market prices could be altered so as to achieve some of these objectives without as much fiscal redistribution? This is exactly what the Swedish welfare state did. The social democratic coalition was built on the corporatist model wherein trade unions and state bureaucracy directly legislated the labor market. This generated higher wages for workers and meant that was less need for redistribution from the owners of capital and corporations to labor. (2019, 475)
This passage is about wage policy and inequality, but the same message would apply to privatization. For as I mentioned above, Heath makes very similar remarks with respect to the choice between markets plus redistribution through tax-and-transfer measures and in-kind provision as public goods, arguing for the economic superiority of the former (2023a, 50). As Acemoglu and Robinson suggest, political dynamics should be weighed in when setting up such markets, and this may lead to different market designs from the ones we would adopt if we merely took an economic lens. In particular, we could look at market designs in which public interests are realized through the actions of private actors themselves, so that governments do not need to correct them post hoc through the fiscal system (ownership design is part of that picture; see the next section). These options are now absent in Heath’s account which contrasts public provision with a textbook picture of competitive markets.

Both observations in this section are meant to show the limits of anodyne theorizing based on economics alone. Intellectually, the anodyne perspective is fine as long as we deal with easy cases, and in this sense is like skating on a lake where the ice is solid and perfectly frozen. But as soon as we approach the places where the ice is thinner and holes may lay hidden, it provides less confident guidance. Practically, we need to know ‘what we buy’ when we accept privatization, and the differences in markets here suggest the need for market typologies addressing different allocations of power, and political-economic analyses of the interaction between politics and the market, in addition to economic theory.

V. THE OWNERSHIP OF PRIVATIZED CORPORATIONS

As we saw (see the quote in section IV), Heath argued that ownership is itself insufficient in determining the outcomes of privatization. Both privately and publicly owned companies have their success stories as well as major failures. In his view, the company’s institutional environment (whether markets are competitive or not) is the pre-eminent consideration. But I want to argue that the internal ownership structure of the company is an equally important determinant in the patterns of control and profits arising after privatization. Hence any government contemplating a privatization must think about the kind of ownership structure it wants to see prevail after privatization.\(^{13}\)

\(^{13}\) Note that I am focusing here on Heath’s type 2, 3, 4, and 5 privatization, not on type 1 where ownership remains public. One can even question whether type 1 cases are instances of privatization since there is no change in ownership structure.
One approach to this question is that once you privatize, the ownership structure is left to private initiative. The assumption is that when ownership is not public, it is private—and while the latter can take different shapes, these different options are beyond the remit of the privatization debate. This is probably the reason Heath does not address the ownership question in his paper, although he deals with it extensively elsewhere in his work (see below). However, the standard ownership structure in the private sector—the investor-owned and shareholder-driven firm—tends to put pressure on the interests of other stakeholders related to the firm. In many instances, these pressures motivate criticisms of privatization. So when thinking about privatization, alternatives for the investor-owned firm need to be on the table (either as mandatory structures for the sector, or as options to facilitate or stimulate), to see whether these worries can be assuaged.

In investor-owned companies, outsider equity investors own the shares instead of one of the other corporate constituencies (such as workers or consumers). These companies are shareholder-driven when the board of the company has a fiduciary duty to shareholders alone, understood normally as the injunction to maximize their profits. This arrangement would be socially optimal for all parties involved. However, lawyers (Stout 2012; Robé 2011; Mayer 2013), political theorists (Ciepley 2013; Singer 2018b) and economists (Mehrotra and Morck 2017; Magill, Quinzii, and Rochet 2015) have argued that shareholder-driven firms tend to exploit non-shareholders. Heath does not deny this risk. He accepts that firms can often “exercise significant market power over its constituency groups and that law and contract typically offer them incomplete protection” (2014, 140). In his view the solution lies in an appeal to ethics: managers should have the “moral imperative […] to avoid taking advantage of the situation”, operate “subject to a set of deontic constraints in dealing with other constituency groups” (141).

The relevant moral imperatives are derived from the market failure approach (MFA) to business ethics. On Heath's account, these offer a clearer and more parsimonious set of norms compared to stakeholder theories of the firm.¹⁴ Heath’s argument for resisting these theories (“the

---

¹⁴ I will leave aside the debate about whether the norms of MFA are sufficient as a benchmark for firms’ ethical obligations, or whether a wider range of moral norms is more appropriate (Singer 2018a; Blunden 2022); and also the debate about whether these norms are more or less equivalent from what would follow from stakeholder theory. Marc Fleurbaey argues that the stakeholder approach, maximizing the total surplus of
deep, well-worn rut of multi-fiduciary stakeholder theory” (2014, 139)) is that this creates inefficient forms of decision-making. Based on Hansmann (1996), Heath argues that the costs of collective decision-making by non-investor patron groups are too high to make this structure efficient (for most industries), because investors have a more homogenous interest (profits) than other patron groups. Investor control creates less conflicts in decision-making, hence lower costs. The proof is counterfactual: otherwise, non-investor groups would already have taken ownership, in a much wider range of markets. This does not mean that non-investor groups are better off under investor ownership; indeed, even society (the sum of all groups) could be better off under non-investor control. However, this position cannot be reached due to the transaction costs for non-investor groups (for details, see Heath 2014, 132–136). Ethics is supposed to make up for the difference, to make managers do without transaction costs what the non-investor groups cannot do for themselves.

To appreciate Heath’s position, I want to frame the issues in terms of the question which form of corporate governance is more effective: an ethical approach relying on managerial moral duties; or a normative approach which gives control rights to stakeholders, so that they can bargain within the corporate governance structure to protect their interests.\footnote{Here I abstract from the otherwise important distinction between control rights (which give voice within corporate governance) and profit rights (the right to receive dividends). One can give non-investors control rights without profit rights (as in German co-determination system), or via giving full ownership rights (as in cooperatives). I take both as possible pathways, under the umbrella of ‘ownership’, in this section.}

This choice in my view does not run parallel to the choice between MFA and stakeholder theory, but is orthogonal to it. Both MFA and stakeholder theory can either defend their norms as (merely) moral obligations for managers or as norms that are to be enforced through control rights for stakeholders.\footnote{Indeed, stakeholder theory is itself divided on the question. Early stakeholder theories advocated giving control rights to stakeholders, but later stakeholder theories refrained from doing so. Stakeholder theory became a theory telling managers how to rule, but not about introducing ‘stakeholder democracy’ (Moriarty 2014). Interestingly, there is now a new wave of stakeholder theories again arguing for stakeholder control rights, based on insights from economic theory (Mehrotra and Morck 2017; Magill, Quinzii, and Rochet 2015; Fleurbaey 2023; Fleurbaey and Ponthière 2023; Stoelhorst and Vishwanathan 2022).}

How to weigh the additional costs of decision-making with a larger set of stakeholders against the benefits of doing so?

On the cost side, it seems to me hard to deny that implementation of MFA and stakeholder theory have higher decision-making costs compared the firm to all stakeholders, yields management rules that are equivalent to Heath’s market failure approach (2023, 249).
to shareholder theory. Stakeholder theory advocates most often accept this. As to the MFA, we need to remind ourselves the norms of market failure are indeterminate, and require quite a bit of interpretative work, or ‘judgment’, to apply them to the real world (Moriarty 2020; Bennett 2023; Singer and Ron 2023). Heath acknowledges this, and calls for ‘phronesis’ by managers in the application of the norms of market failure (2023b, 103). But once this is acknowledged, it seems to me that the force of the traditional objection against stakeholder theories—that they do not provide a single objective function to evaluate managerial success—is considerably weakened. That point, repeated over and over by shareholder-primacy theorists, depends on the idea that stakeholders take care of their own interests through contracting with the firm. Heath (rightly) rejects this “let them eat contracts” view (2014, 136). But while the MFA avoids the language of fiduciary duties to stakeholders, the moral duties to stakeholders that enter via the categories of market failure still do need to be balanced with the fiduciary duties to shareholders. Hence the decision-making costs of the MFA-abiding firm are also higher than those of the shareholder-wealth maximizing firm. Managers need to find out what the moral imperatives or deontic constraints are, which apply to their decision-making situations. A heterogeneity of interests enters the calculus. To get information about the size of negative externalities, or the extent to which stakeholders suffer from information asymmetries, they may need to be consulted. At the end of the day, a balancing of considerations must take place. These activities, associated with stakeholder management, will also be needed when managers follow the MFA.

Based on this acceptance of the enhanced costs of any ethical approach, we need an estimation of the additional increase in decision-making costs when moving from the ethical to the normative approach. Additionally, that is, to the increase in moving from the shareholder wealth maximizing approach (made famous by Milton Friedman 1970) to the ethical approach (MFA or stakeholder theory). If, as I suggested, the real cost increases are in the latter move, then the cost increases in the additional move to political empowerment will probably be relatively modest. The information and consultation sessions with stakeholders now change into bargaining sessions, but they should take the same amount of hours (if upon a shift from the ethical to the normative approach, one discovers that the latter take many more hours, then one should wonder whether stakeholder interests had effectively been protected with the information and consultation sessions in the first place).
On the benefit side, it would seem that a normative approach offers more secure enforcement of legitimate stakeholder interests that cannot be effectively enforced through law or contract. For, other things equal, a norm backed-up by real bargaining power is more effective than a norm which must rely on the goodwill of the person who has to sacrifice to adhere to that norm (for a full account in the context of stakeholder theory, see the arguments by Moriarty 2014). This is why important moral norms (such as prohibitions of criminal acts) get converted into politically and legally enforced norms, despite their costs of enforcement. This is one way of putting the issue: when making the same costs, the ethical approach leads to a lower level of benefits (stakeholder protection or MFA-norm-adherence) than the normative approach. Alternatively, we could hold the level of benefits in our comparison constant, for after all, these benefits are not optional: they represent moral norms. So we want to imagine a situation in which both approaches are effective in realizing adherence to these norms.\textsuperscript{17} It seems that decision-making costs are only one factor. A much more important factor is that under the ethical approach, the firm will often face a competitive disadvantage against firms that do not behave as ethically (as Fleurbaey 2023, 249 says: “responsible management is as likely to win the market competition as a clean cyclist is likely to win the Tour de France”). By levelling the playing field, a normative approach solves that problem. It is quite hard to imagine how the ethical approach could lead to the \textit{same} level of stakeholder protection (MFA-norm-adherence), against these costs.\textsuperscript{18}

\textsuperscript{17} This makes the cost-benefit analysis here different from Hansmann’s approach. The latter is based on a bargaining process of what benefits each constituency group is willing and able to bid for, to assume ownership (hence assuming their initial endowments are just), not on the norms of separate moral theory of what benefit-level they are entitled to.

\textsuperscript{18} The ethical approach would need to make huge costs to overcome the competitive disadvantage. Call these \textit{advocacy costs}. In the current cultural context (social norms) \textit{A}, the effectiveness of reliance on managers’ moral conscience (possibly supplemented with the effects of social shaming by non-firm actors on managers) leads to norm-adherence at level \textit{a}. We cannot expect any further headway to the desirable (or morally required) level \textit{b}, at least without additional advocacy costs (for example, the financial, time-related, and reputational costs of collective actions by academics, NGOs, and conscientious business leaders in pressuring businesses for following their moral compasses). Only eating these costs will (maybe) shift society towards a new cultural context \textit{B} (set of social norms), in which firm managers are effectively compelled to shift the boundary of what their firms are prepared to do to \textit{b} (against this, one might object that shifting a society to \textit{B} is also necessary to create the political will to legislate mandatory stakeholder rights; and hence the same advocacy costs apply. That may be right. But in \textit{B}, as in \textit{A}, the political approach will lead to higher norm-adherence than the ethical approach).
Heath agrees that law and contract leave stakeholders with insufficient protection: there is an ‘enforcement gap’ after law and contract have been optimized. But we must go beyond declaring (as philosophers are good at doing, *ex cathedra*) what the relevant moral norms are which need to bridge the gap. We need to think about what is needed to actually get there. In this context, firm-level protection of stakeholder interests is best seen as an intermediary between ‘perfect centralization’ of decision-making, at the level of the state (law), and ‘perfect decentralization’, at the level of the market (contract) (Stoelhorst and Vishwanathan 2022, 12). Both extremes have been theoretical starting points for long-standing traditions of thought, respectively socialist state planning and *laissez-faire* economics. It is unthinkable that one of these levels on its own will be best suited to solve *all* problems of social coordination. Hence, we will always deal with a mix of law and contract, and we can pragmatically think about the distribution of moral labor between both levels as a matter of cost-benefit analysis. But why exclude the vast terrain in between? Here decisions are not centralized to the state, nor left to individuals acting on the market, but centralized for a certain subset of individuals smaller than ‘all citizens’. The firm is the primary example. From the same pragmatic point of view, it would be surprising if this intermediate level would never be the best option to solve social coordination issues, and all of them would be best dealt with at the extremes of state centralization and contractual decentralization.

It may seem we have made quite a roundabout with respect to the issue of privatization. But ownership design is part of market design, and hence very relevant to privatization. When privatizing, governments can decide to worry about market design (the topic of the previous section) but decide to leave whatever ownership forms emerge to spontaneous processes. But alternatively, they can encourage or even mandate particular governance and ownership structures to help realize public interests that cannot effectively regulated by law or contract. Benefit corporations, cooperatives, foundation-owned companies, or social enterprises, are some examples of these. They have proven economically viable and successful in various countries and sectors, as research on cooperatives (Mallowson 2014) and foundation-owned companies (Hansmann and Thomsen 2021) shows. The absence of this theme from Heath’s discussion leaves us with a dichotomy, in which firms are either publicly owned or privately owned by their shareholders (or mixes between these two, as when discussing State-Owned Enterprises in 2023a, 55).
Ironically, this dichotomy may make privatization less attractive to those who are reluctant to introduce market competition. Ownership structures which provide more protections for stakeholder interests can be a promising avenue to combine the virtues of decentralized decision-making and market competition with these protections. This may win over some who would otherwise refuse to even think about privatization.

VI. CONCLUSION
Heath has done philosophers a tremendous service in enriching the debate on privatization with his economics-inspired view. His paper provides lots of analytical tools which helps to look at privatizations on a case-by-case basis and in a pragmatic spirit. With his approach, he positions himself on the other end of a dichotomy of non-instrumentalist (deontological) approaches on the one hand and instrumentalist-welfarist approaches on the other hand. My aim in this paper was to search for an alternative to break through the dichotomy. I have done so by offering a broadly instrumentalist theory which is nonetheless not welfarist. In my view, the ends we seek when thinking about socially vital goods (our theory of public interests) should go beyond Pareto-efficiency. As to the means we employ to realize these ends, we need a view of markets which takes into account not just their competitiveness, but also the distribution of power. This means we need to differentiate market types. Also, we need to differentiate ownership types. On both counts, I have found the picture Heath sketches leaves out too many of the variables that, in the end, may be decisive in whether or not a privatization is acceptable. When considering to privatize a public service, the central task is to think hard about what distribution of power and benefits we will get when we do so—including the question of how confident we can be that the government can design markets according to its plans and control the ‘animal spirits’ of today’s financialized capitalism. Heath debunks the all-too-easy reluctance to privatize that some may still have. Moving forward, the economic frameworks on which he relies would best be employed in combination with the theoretical resources discussed in my response.

REFERENCES


**Rutger Claassen** is Professor of Political Philosophy and Economic Ethics, at the Ethics Institute of Utrecht University. For more information, please visit [www.rutgerclaassen.nl](http://www.rutgerclaassen.nl)

Contact e-mail: <R.j.g.claassen@uu.nl>
Public Provision in Democratic Societies: Reasons to Reject Privatization

MARTIN O’NEILL
University of York

Abstract: If we hope to see values of equality and democracy embodied in our societies’ institutions, then we have a range of good reasons to favor expansive public provision of goods and services, and to oppose many forms of privatization. While Joseph Heath is right to argue that there are at least some forms of ‘anodyne privatization’, and while he is also right to argue for a more nuanced philosophical debate about the different dimensions of choice between forms of public and private provision, Heath fails to register various regards in which private provision can undermine these central public values. We often have strong egalitarian and democratic reasons to protect zones of decommodification; to resist the imposition of user-charges; and to favor insourcing and direct public procurement over various forms of outsourcing of public services. Public libraries provide a totemic illustration of some of the deep virtues of collective public provision in democratic societies. Overall, our reasons to reject privatization are stronger and more diverse than theorists such as Heath might have supposed.

Keywords: privatization, public goods, public provision, public sector, outsourcing, social justice

JEL Classification: D4, H4, H6, L3, P16

I. INTRODUCTION

Joseph Heath’s article on “Anodyne Privatization” contains a great deal with which even those who are in general opposed to privatization can agree. Heath is right to highlight the diversity and heterogeneity of the different kinds of institutional and administrative changes that can fall...
under the umbrella of ‘privatization’, and his discussion provides a helpful typology of these different kinds of privatization, which will be of significant value as the debate on these issues becomes more developed and more nuanced. Heath is also right that the recent debate on privatization, with its (understandable) focus on some of the more outrageous and objectionable abrogations of the state’s core functions in favor of the activity of private agencies—as in the privatization of prisons and border security, and with the enlarged role for private military contractors in many parts of the world (Cordelli 2020, 2021; Dorfman and Harel 2013, 2016, 2021; Vasanthakumar 2021)—can sometimes seem to have occluded discussion of more humdrum and quotidian forms of privatization, especially those occurring further from the state’s fundamental core functions in providing law enforcement, defense and security. Heath is right, too, that we should not be too quickly led from conclusions in one area of policy to more general conclusions that may on reflection fail to take seriously the diversity of the issues at stake when we consider which functions should be exercised by the state and its employees, and which might legitimately be performed by private agencies.

As I shall explore in more detail below, I am also highly sympathetic to one important element of Heath’s view, in that I share his sense that discussion of public provision has been confused and led astray by the treatment of these issues within economics, and especially by the canonical treatment of public goods that we owe to Paul Samuelson, according to which public goods are marked out essentially by their indivisibility and non-excludability. As Heath agrees, many goods and services that we have good reason to see provided by the state do not in fact bear these canonical features of classic Samuelsonian ‘public goods’. Hence we need in this area to have a discussion that is not excessively constrained by the influence of Samuelson’s particular way of thinking about the issues. But while I am happy to applaud Heath for his readiness to twist free of the misleading framing given to us by some parts of standard textbook economics, my central critique of Heath’s approach to public provision is that he remains very much in the grip of an economistic picture of the central issues that are at stake when we think about public services and public spending. Heath’s way of thinking about the reasons for public provision turn on the central value of efficiency as a good to be pursued in public policy, and he looks upon state provision as justifiable only when it is an efficiency-enhancing response to various forms of market failure. While I do not deny that Heath often identifies good reasons for
removing certain goods or services from the market, his ‘market failures’ approach (Heath 2011) is far from exhaustive and, as I shall argue, misses several strong reasons we may have for providing goods and services via the state and its agencies, outside of the market. In democratic societies, efficiency is, and should be, an important value, but there are broader normative reasons for being against privatization in many relevant cases. Therefore, as we shall see, many of Heath’s cases of ‘anodyne privatization’ are not in fact ‘anodyne’ at all. We often have good reasons to resist privatization, in its different forms, that are not captured by Heath’s approach.

This article will have the following structure. Section II discusses Heath’s presentation of the ‘Traditional Public Sector Model’ (TPSM), and his account of the bad influence of Samuelson’s account of public goods. Section III takes issue with Heath’s critique of the idea of decommodification, in light of a more compelling and expansive conception of egalitarianism. Section IV discusses Heath’s (useful and clarifying) model of the different types of privatization (roughly speaking, the distinction between ‘supply side’ and ‘demand side’ privatization), in light of cases such as the introduction of higher education charges and the ‘contracting out’ of some local public services such as rubbish collection and waste management. Section V takes issue with Heath’s ‘free lunch’ critique of arguments against any role for profit-seeking firms in public service provision. Section VI suggests that we have good reason to reject Heath’s account of some purportedly uncontroversial cases of ‘type-1’ privatization (for example, the introduction of user charges), while section VII makes the case for the inadequacy of Heath’s account of ‘type-2’ privatization (such as ‘contracting out’ of certain state services), and argues that we often have good reasons for insourcing and direct public procurement. Section VIII gives a more sympathetic comment on Heath’s discussion of ‘partial’ privatization and the distorting effects of public accounting rules. Section IX is largely concerned with one of the greatest of all public institutions: public libraries (along with other institutions that provide non-commercial public spaces). Section X concludes, drawing the foregoing elements together to show the inadequacy of Heath’s approach, and showing why those of us who hope to see values of equality and democracy enshrined in the institutions of our societies have more reasons to be against privatization—and in favor of expansive public provision of goods and services—than Heath would have led us to believe.
II. THE TRADITIONAL PUBLIC SECTOR MODEL, AND THE BEWIN CHING EFFECT OF THE STANDARD ECONOMISTS’ ACCOUNT OF PUBLIC GOODS

Let us begin with some scene-setting. Heath helpfully brings in the idea of the ‘Traditional Public Sector Model’ (TPSM), as described by the economist Evan Davis (Davis 1998). As Heath puts it:

The key characteristic of this model is that it involves the provision of some good by employees of the state, funded through general taxation, such that it is free to citizens at the point of service. [...] The crucial feature of this model is the fact that the state stands on both sides of the transaction—as both the purchaser (that is, the source of funding) and the provider. (2023, 31–32)

One only has to see this description of the TPSM to realize that there are already (at least) two questions regarding public provision and public services, and two domains in which we might think about (different forms of) privatization: the question of who provides the good or service (that is to say, by state employees, arms-length agencies, quangos, private companies, etc.); and the question of how the good or service is funded (through general taxation, hypothecated taxation, by regulated user fees, by some mix of these elements, or simply by individuals paying a market price for the service, etc.). As Heath rightly points out, there is something about the particular example of the British National Health Service (NHS) that draws our attention towards it as an exemplar of quite a pure form of the TPSM, with state employees providing a service to citizens, funded through general taxation, and free at the point of use. (Although even here matters are more complicated than they would be on a fully ‘pure’ version of the TPSM, with many citizens (but not children, pensioners, or those in receipt of certain public benefits) needing to pay charges for eyecare, dental care, and for prescription drugs). But as Heath points out, the NHS is in some respects quite unusual, insofar as in many other countries with de facto universal healthcare provision, this is achieved not via the TPSM, but through regulated insurance systems characterized by careful state intervention, where a great deal of service provision is by private providers.

1 Aneurin ‘Nye’ Bevan, the government minister who led the introduction of Britain’s NHS, resigned from Clement Attlee’s government in 1951, in protest at the introduction of charges for eye tests and dental treatment. See Morgan (1985); Foot (2009); and Thomas-Symonds (2014).
As Heath also rightly reminds us, even when we think of a socialist political party such as the British Labour Party under the leadership of Clement Attlee, when it still had a principled commitment to the nationalization of the ‘commanding heights’ of the economy (Morgan 1985), such as civil aviation, electricity and gas production, the coal and steel industry, and the railways, the idea was never that the products of these industries be made available to all free at the point of use. In one way this is a point so clear as hardly to need stating, but it does at least have the virtue of drawing our attention to the split between separable questions of funding, control, availability, and cost, that can often unhelpfully be merged under the single heading of privatization or nationalization.

Given that the TPSM is in some ways rare, and “given the ubiquity of purchaser-provider splits in the public sector” (Heath 2023, 33) it does seem puzzling that, notwithstanding the lure of the powerful example of the NHS, debate on these issues has so often lacked nuance about these kinds of splits and mixed cases. Heath’s diagnosis lays at least part of the blame on the influence of Samuelson’s treatment of ‘public goods’ in his famous and widely-influential textbook—Samuelson and Nordhaus 2009—according to which public goods are “perfectly non-rival and non-excludable” (Heath 2023, 33). For true Samuelsonian public goods, such as national defense or environmental regulation, only the TPSM generally makes sense as a method of provision. And so, while in the grip of Samuelson’s picture of public goods, we are led away from a more nuanced engagement with the normative issues at stake. Here Heath reaches a conclusion similar to one I have defended in earlier work (O’Neill 2020b, esp. 180–190), where I made the case that John Rawls’ discussion of the relationship between the public and private sectors was badly distorted by his adoption of this ‘economistic’ conception of public goods (with, in Rawls’s case, the influence being in particular via the work of James Buchanan (see Rawls 1999, esp. 235–239; Buchanan 1968). My worry here was that:

Rawls seems to slide from the acceptance of the economists’ conception of public goods to a normative position according to which it is only when a good meets this strict definition that there is a

---

1 Samuelson’s hugely influential textbook was first published in 1948, and went through 19 editions over the next 60 years. The first 12 editions were written by Samuelson alone, with later editions (up to the final one in 2009) co-written with William D. Nordhaus. The book has sold something over 4 million copies overall, and has been translated into over forty languages. On the book’s history see Skousen (1997).
justification in terms of justice for making its distribution a matter of public democratic decision making rather than a matter for the market. (O’Neill 2020b, 185)

Heath’s article has convinced me that the baleful influence of this particular unwarranted conceptual slide has actually extended rather more broadly within political philosophy. My concern in that earlier piece had been that, when it comes to public goods and public provision, “an overly stark and schematic presentation of the conceptual terrain again makes it more difficult to situate some of the most important issues around the role of the state in a just society” (O’Neill 2020b, 186), and so it is far from faint praise to say that Heath’s article does sterling service in rescuing its readers from the temptations of that overly stark and schematic picture.³

I have said that my praise for Heath is certainly not faint, but it is limited in scope, nonetheless. For after rightly diagnosing this particular instance of what we might think of as ‘the cognitive harm of economism’ Heath immediately commits a rather similar kind of error. Heath tells us, plausibly, that “a non-excludable good is just the limit case of a good that generates positive externalities” (2023, 33), but this leads him to slip towards a further and less plausible claim. Heath suggests that, just as it might make sense to go with TPSM-type provision in just these limit cases, we should also think that, where there is some measurable ratio between private benefit and public benefit, that the split between private and public funding of that good should be strictly proportional to that division of benefit. As Heath puts it, “the appropriate price to charge for such a good is dictated by the fraction of the good that constitutes private value (that is, appropriated by the individual), and the fraction that takes the form of a positive externality” (33). Unfortunately, this really is a non sequitur, or at least that is so in the absence of any independent normative principle that explains why the cost of state services to citizens should be strictly proportional to the benefit they receive from those services. But no such principle is advanced, and neither is an argument offered that might support such a principle.

When we talk about the question of how state services are funded, and how the state should allocate benefits from public spending, we are

---

³ In that earlier discussion, I took public libraries as my paradigm case of an important public institution (and part of the ‘TPSM’) which provides a good that is neither ‘perfectly non-rival’ nor ‘non-excludable’. I shall return to this case in light of Heath’s own arguments in Section IX below.
precisely in the business of arguing about the question of when and how individuals might benefit in a non-proportional way from the activities of the state, which they might fund (proportionately or non-proportionately) either through their taxes or through direct or indirect user fees. This is a difficult normative question which cannot be settled by the illicit importation of the tacit assumption that what lies behind state provision of goods or services is just the concern to provide benefits, in an efficient manner, in a way that spreads costs in proportion to benefit received. The unexamined view in the background here seems to be along the same lines as the thinking that leads to the quasi-libertarian idea of a ‘benefits tax’—“that is, a form of taxation that looks to set rates of taxation in a way designed to mirror market outcomes, by taxing individuals for their consumption of public goods at a level that as closely as possible parallels the market prices of such goods” (see O’Neill & Orr 2018, 10). Given the lack of a market for many such goods, such a proposal would rely on positing ‘shadow’ market prices for these various goods and services, and the computation of these ‘shadow prices’ would in turn depend on a number of conceptually and normatively contestable assumptions, not least of which would be assumptions regarding the structure of the imagined shadow market that would be needed in order to yield determinate prices (10). I realize of course that there is much that could be said about such a view advocating proportionate costs and benefits for public goods and services, either in defense or in critique, but such a view cannot be allowed to pass along simply by means of an unexamined and (only superficially plausible) argumentative sleight of hand.⁴

III. Decommodification and Social Equality
Let me now turn to what Heath has to say about the idea of decommodification, about which he is excessively skeptical. On Heath’s view, another reason why our attention has been overly directed at the TPSM is that too much weight has been given to the idea “that an important function of the welfare state is to achieve a ‘decommodification’ of certain goods” (2023, 34), as per the ambition stated in the NHS constitution that access to healthcare services should be “based on clinical need, not an individual’s ability to pay” (DHSC 2023). This idea of decommodification naturally leads us towards wishing to see the expulsion of market pressures and market mechanisms from the provision of these important goods and

⁴ For what I take to be a magisterial examination of the idea of a ‘Benefits Tax’, see the discussion by Barbara Fried (2018).
services; but Heath thinks that this line of thought frequently rests on various kinds of mistakes. For one, he claims that even with the NHS, healthcare is distributed not in accord with healthcare needs per se but “in accordance with a physician’s determination of a patient’s needs” (2023, 34). But this is clearly a rather shallow point: after all, physicians’ determinations of need are intended to track underlying need, and will frequently do a much better job of tracking medical need than will the beliefs or preferences of the patient themselves. This is not to say that doctors cannot be mistaken, of course, but it is to say that the architects of the NHS certainly did not think that they were building an institution that could somehow work without the exercise of expert judgement. It should be no part of the normative ideal of decommodification to hold that we should be aiming for a society that distributes those decommodified goods without any administrative intermediation. There is also here of course the point familiar from the work of T. M. Scanlon (1975) and others that individuals’ preferences in respect of features in light of which they may have a claim on public support may not align well with objective aspects of those features: in other words, preference and urgency can come apart. A patient may prefer cosmetic dentistry to give them a more handsome smile, but what they might actually need could be a root canal treatment to save an imperiled molar.

A more weighty criticism that Heath advances against the idea of decommodification is that egalitarians are making a kind of conceptual mistake when they think that decommodification advances the goal of equality, or when they think that it is mandated by our concern that the state treats its citizens as equals. Heath sees this as amounting to something of an unstable mixture of (mere) “distaste for the commodity form” (2023, 35) and a mistaken belief that decommodification is the best way towards achieving more egalitarian distributive outcomes. Heath seems to find this first element simply mystifying, while arguing that decommodification is unlikely to be supported by egalitarian reasons “because the same objective could be achieved by redistributing income, or by providing individuals with vouchers to access certain goods” (35). Here we have Heath’s rather economistic outlook leaving him unable to make good sense of what is actually a readily intelligible, indeed compelling, worry about the anti-egalitarian social consequences of marketization.

Let me try to explain why with an analogy. Imagine I have four children, whom I love each equally, and whom I am equally motivated to help to flourish. With these motivations, my aim may be to give their needs
equal weight in my deliberations about how to act with and towards them, and my hope would be that their understanding of my behavior would be that it embodied and expressed this kind of equal love and concern. Now imagine a different parent, acting in the grip of their exposure to economic (or, perhaps, economistic) ideas. Such a parent might come to the view that their best way to treat their children equally would be to figure out their overall budget of time, money, and attention, and to give each of their children a set of vouchers that they could use to make claims against them. Such a system would presumably have some of the general advantages of efficiency associated with free markets. On further reflection, this system itself might come to seem too restrictive, and could be replaced with a more radical version whereby all parents issued such vouchers, and children could use their vouchers to make claims on anyone operating within the general voucher market. The returns to scale would presumably be tremendous. And yet I do not think it would be eccentric or esoteric to suggest that something of great value would be lost in the introduction of market mechanisms here. The parent-child relationship has value in itself, and the value of that relationship can be expressed directly by the way its participants stand towards each other; we care not just about ‘who gets what’ but also about the character of the relationships that play out day to day in the course of people living together.

Now, I can imagine Heath interjecting—so what? State provision of public services does not involve the kinds of relationships we see within families. Well, that’s true of course. But it does involve certain kinds of relationships—of individuals to the state, and of those individuals to one another—that are not identical to market relationships. To take another not-unrelated example, when my (actual, as opposed to hypothetical) twins were born by Caesarean section in an NHS hospital, there were quite a few highly trained people in the room—doctors, nurses, auxiliary staff. They were there because they were all needed for my children to enter the world safely and healthily, and for their mother to likewise get through the experience safely and securely. Their presence there as employees of a public service instantiated and embodied a commitment to a certain general conception of serving and benefitting the public, embodied in this case as the actual patients there in need of their care and
expertise. A society expresses something when it elevates these interactions between people, and takes them outside the ambit of the market. My twins came into the world making a claim on the attention and hard work of others both as citizens and, more generally, simply as individuals in need of care; not as market participants, or as individuals on whose behalf other market participants were acting. It is perhaps true to say that the right arrangement of income-contingent vouchers, or whatever it might be, could have achieved a more egalitarian outcome in sheer distributive terms, but it could not have achieved an instance that embodied a group of people relating to one another, outside the market, as participants in an institution whose very foundations were about an egalitarian expression of the value, standing and status of all participants within society.

This brings us back to Heath’s dismissal of “distaste for the commodity form” (35), which makes it seem as if opposition to so many human interactions being mediated by commodity exchange were no more than a raw sense of displeasure, or an unelaborated sense of ‘ick’. But this ‘distaste’ can have much deeper roots, and can itself be seen as resulting from a broader conception of the value of relating to one another as equals. If, as I do, one sees the value of equality as tied in to ideas of equal standing and status, and to a certain ideal of how people might relate to each other, it is not difficult to see how centering the value of equality can give us reason to prefer forms of interaction in which we are participants in shared institutions that themselves affirm our equal status and value, as opposed to forms of interaction where we face each other only as market participants, each out to get what they can for themselves as efficiently as possible (O’Neill 2008, 2013, 2016).

Heath claims that “the egalitarian perspective offers no explanation for why the state needs to be more deeply involved on either the purchaser or the provider side” (2023, 35), but this is true only on a rather impoverished and attenuated understanding of what ‘the egalitarian perspective’ really is. If we move beyond a limited notion of equality as being no more than a merely distributive ideal, in favor of the richer and more substantive ideal of social or relational egalitarianism that I (among many others) have explored in some detail elsewhere, then we quickly see that the egalitarian perspective is precisely a perspective from which we have

---

5 As Michael Walzer puts it in *Spheres of Justice*, workers within a national health service “serve for the sake of the social need and not, or not simply, for their own sakes” (1983, 89).
good reason to want to limit the scope of the market, and the centrality of the commodity form. If we want to value and promote egalitarian social relations (see Fourie et al. 2015, Scanlon 2018, Schemmel 2021, among others) then we instead have reason to endorse institutional forms that embody and express the value of equality through giving people the opportunity to act with others in a way that brings them outside the self-interest or even antagonism of market relations. On this kind of social egalitarian view, it is not (pace Heath) any kind of a mistake to favor decommodification of certain especially significant goods and services.

IV. MODELS OF PRIVATIZATION, HIGHER EDUCATION, AND CONTRACTING OUT

Let me return now to some praise again, and in particular to my appreciation of Heath’s typology of the different varieties of privatization. By thinking about the contrast between the provision and purchase of different goods and services, we can clarify different processes that might all be described as forms of ‘privatization’. Consider for example the difference between, say, (a) the British state selling-off British Airways or British Telecom (wherein we shift from state provision and private purchase to private provision and private purchase), and to return to the example that drove Nye Bevan to resign his ministerial position, (b) the case where state funding for eyecare was withdrawn (wherein we shift from state purchase and private provision to private purchase and provision). As Heath rightly points out, perhaps the starkest form of privatization has been where we have seen a shift from a TPSM model of residential water provision to something much more like a fully private market (2023, 38), as in the case of the Thatcher government’s privatization of the water industry in the late 1980s (see Parker 2012). An interesting contrast here would be the Irish case, where pressure from the EU ‘Troika’ after the financial crisis brought the Irish government in 2013 to introduce water charges, in a country where residential water services had previously been provided by the state free of charge. The water company created during these changes, Irish Water (later Uisce Éireann), remains a state-owned utility, and so here one saw a transition (if only temporarily) from ‘decommodified’ public provision to a form of private purchase that coexisted with provision by a state-owned utility (see Brennan 2019). All these different cases are distinct in ways that Heath’s typology helps to illustrate; and each raises its own distinctive kinds of normative issues.
Heath is right to say that it is a great source of confusion that the term ‘privatization’ can be used variously to describe not only transitions from the TPSM to a fully private market ('FPM' in his terms), but also to describe the very different phenomena of transitions from the TPSM to a ‘supply-side public sector model' (SPSM), for example as with Irish Water, or a ‘demand-side public sector model' (DPSM), for example when private providers are introduced into the NHS even though care remains free at the point of use (2023, 39). Where I am less inclined to follow Heath, though, is in his claim that such ‘one-sided' privatizations are very often ‘anodyne', "simply because the state retains the ability to exercise considerable control over the market by virtue of dominating one side of it" (39). Notwithstanding this point about the potential for residual forms of state control, Heath's approach is severely liable to underestimate the harms that can be done both by the 'commodification' (or 'recommodification') involved in the introduction of user charges for what had once been generally accessible public goods or services; and the harms attendant on 'outsourcing' certain aspects of public provision to private firms.

One interesting case is higher education. Heath is rather relaxed about the introduction of variable fees in higher education, seeing this as just a way of reallocating costs within a system that can nevertheless remain largely under the regulation of the state. But the changes in higher education that have been seen in many countries in recent years are not just about a redistribution of costs with respect to an otherwise unchanged system. Rather, the introduction of high and/or variable fees is generally accompanied by a systematic change of culture in higher education, including a greater tendency towards the ‘instrumentalization' of learning and education; a funneling of students towards courses that are more straightforwardly remunerative in economic terms, and away from subjects and courses whose value may be best captured in non-pecuniary ways; and a transformation of self-conception among students, academics and (especially) among the administrators and managers of academic institutions, who come to see the enterprise in which they are engaged in more atomized and competitive terms, and less as a domain of cooperation and collaboration. Many writers have spoken here in terms of the promotion of a 'neoliberal' mindset, that transforms the self-understanding and the social relations of the participants in those institutions (Fisher 2009; Davies 2014; Newfield 2021; Boyle, Hickson and Ujhelyi Gomez 2023). One might say that these changes are perhaps not a necessary feature of any such change in the distribution of costs, but for those of us
thinking about the organization of institutions of higher education ‘around here and around now’, so to speak, these broader changes are real and significant, and should be taken with sufficient seriousness.

Another very different kind of example would be with the now widespread practice of city councils and other local authorities ‘contracting out’ services such as cleaning, maintenance, and dealing with domestic and business waste and sanitation. Heath is right to say that this kind of ‘contracting out’ is often opposed on social justice grounds due to reasons of job quality and distributive justice, given that “government employees are often better paid and enjoy better working conditions, including more extensive unionization, than private-sector employees” (2023, 39). The point here about working conditions certainly bears emphasis, as in general the only reason that contracted-out services can appear as a good deal to local authorities, or to other institutions such as universities or hospitals, is precisely because they will often involve a serious reduction in pay, control over working hours, and job security for those working in those occupations. To this we can add that there are social egalitarian concerns that come into play when those working in the same spatial location find themselves divided into a ‘two-tier’ workforce, wherein those with more highly respected jobs (such as council bureaucrats in local government, academics and senior administrators in universities, medical staff in hospitals, etc.) remain direct employees of the main institution, while those undertaking the typically less well-paid roles that are ‘contracted-out’ find themselves as no longer the colleagues of their more advantaged co-workers, but as no more than causal, insecure agency workers. This obviously is likely to distort relations among employees, and to be injurious to the self-respect of those who are already disadvantaged in other respects. One might with accuracy say that this will often involve adding insult to injury.

In these cases, though, Heath would presumably say that the connection between some of these normative problems of injustice to privatization itself (in its different forms) is in some ways a weak and contingent one. Moreover, there will often be, at least in theory, other remedies available that could counteract some of these problems when and where other reasons—such as reasons of efficiency—point towards the case for privatization. Of course, it is true that these are not necessary connections, but I think it behooves us as political philosophers looking to operate under broadly realistic conditions, and looking to make arguments that will find a foothold in the world around us, that we should take these
common and significant connections seriously. It may not be strictly unavoidable that marketization in universities erodes much of value in their internal institutional culture; or that contracted-out services generally involve a shift towards more demeaning and unpleasant forms of employment for those who are already disadvantaged. Nevertheless, where these connections are common and striking, we should take them seriously. I therefore do not disagree with Heath when he says that “there can be no categorical rejection” (39) of certain forms of either supply-side or demand-side privatization, but I nevertheless hold that it is important for us to realize that the balance of considerations quite often gives us strong reasons to reject privatizations of both kinds. My hope is that something of this case has been made already, and it will hopefully be strengthened by the argument of the remaining sections.

V. PROFIT, EFFICIENCY, AND PUBLIC SERVICES: A TALE OF TWO FREE LUNCHES?

Before turning to some of the further concerns that come up even with cases of ‘supply-side’ and ‘demand-side’ privatization that Heath finds quite ‘anodyne’, I want first to examine Heath’s rather quick dismissal of what he terms (pejoratively) a couple of ‘free lunch’ arguments. The first ‘free lunch’ argument that Heath quickly dismisses is an argument in favor of privatization that holds that “government is necessarily less efficient or less innovative than private enterprise” (40). Such a view seems to be commonly expressed by certain right-wing politicians (most often those whose parties are well-funded by corporate interests), and of course are the mainspring of commentary in ‘pro-business’ publications such as *The Economist* (Zevin 2019). But as Heath rightly allows, no such general claim is remotely plausible, and history is littered with examples of disastrous privatizations which, for from enhancing efficiency or innovation, have simply allowed the appropriation of public wealth by private interests.⁶

---

Heath wants just as quickly to dismiss another purported ‘free lunch’ argument, on the opposite side of the debate, according to which “the public sector is always more efficient because it is not burdened by the expense of paying out ‘profits’ to investors” (2023, 40). Setting aside the curiosity of ‘profits’ being put in what seem to be scare quotes here, Heath is being rather too quick to look for a strict equivalence of bad reasoning. Whereas the relative efficiency or innovativeness of the public and private sectors seems simply to be an open question, with different answers in different contexts (Mazzucato 2013), the question of whether public services should be run only for the benefit of their users and direct beneficiaries, or whether they need also bear the cost of carrying a return for corporate shareholders (not to mention typically highly-remunerated private sector senior managers), is a question amenable to a rather more systematic answer.

Heath’s way of looking to demonstrate that this position is a fallacious ‘free lunch’ is ingenious, but unconvincing. His suggestion is that:

Since every welfare state in the world today runs a substantial budget deficit, every marginal increase in expenditure is financed by debt. And so either way, providers of capital get paid, whether in the form of interest on government bonds to finance public-sector investment, or in the form of profits on shares to finance private-sector investment. (2023, 40)

One response is to point out that it is not true that all states are in budget deficit. At the time of writing (Autumn 2023), Ireland is projecting an annual budget surplus of around €12 billion. The United Kingdom ran a budget surplus for a number of years in the late 1990s and early 2000s, as did the United States around the turn of the century, while Germany ran budget surpluses for the eight years from 2012 to 2019 (inclusive). So even by the lights of Heath’s own argument, this is starting to look like a merely ‘contingent’ free lunch. And this is even before we address the point that any state currently in deficit could always raise taxes to pay for an increase in public provision, such that any expansion of public investment could always be accompanied by fiscal changes to give an overall result that was neutral with regard to the size of the deficit.

---

7 For UK figures, see https://commonslibrary.parliament.uk/research-briefings/sn06167/ For US figures, see https://fiscaldata.treasury.gov/americas-finance-guide/national-deficit/. For German figures, see https://www.worlddata.info/europe/germany/debt.php
Heath’s picture is misleading in other ways though, given that, more significantly, when we think about the privatization of service provision, we are often talking about the operation of local councils, and of local and city governments, who are often barred from debt-financed spending, but must fund all of their spending from current revenues. In those cases, Heath’s supposed equivalence of public debt financing and providing support for private profits is simply irrelevant. Moreover, for Heath’s argument to get a foothold, we would have to imagine circumstances where a public authority of some kind is facing a simple choice between: (a) buying goods or services from a private provider which covers all investment, and bears all investment-related risk; and (b) borrowing the funds itself that are needed to fund that investment and to undertake the associated public provision. But in many actually-existing cases where private providers are brought in, the state (or local council) has already made much of the relevant capital investment, and simply *unavoidably* bears much of the attendant risk, given that the good or service in question will often be something which it is obliged to provide (whether that is sanitation services, transportation, or whatever else). Much private investment in public services exists in a setting of (inevitable) ‘state derisking’, where the profits and other private benefits (for example, in high managerial salaries, etc.) extracted by the private investor themselves amount to what could best be described (in Heath’s own terms) as another kind of ‘free lunch’.8

Let me illustrate with a quick example. When I was growing up, I lived in the London Borough of Ealing. In the 1990s, Ealing Council decided to save money on its rubbish-collection, maintenance, and parks department by bringing in a private provider, which could provide the services at a lower unit cost. The provider chosen was a newly constituted company called BRETS (‘Branch and Root Ealing Technical Services’), which was a subsidiary of US conglomerate Halliburton, later famous for its close political connections to the Republican Party (not least via its sometime CEO, Dick Cheney, who ran the company from 1995–2000), and its profiteering in the wake of the Iraq War.9 Needless to say, the jobs of those who found themselves shifted from council employment to BRETS worsened, service provision did not improve, and the experiment was generally held to be a

---

8 On state derisking and options for the socialisation of risks and rewards, see Mazzucato (2013) and Mazzucato and Li (2020).

9 See information from the Public Services International Research Unit (PSIRU) at the University of Greenwich, here: https://www.psiru.org/companies/halliburton.html. For an entertaining and informative fictionalised account of Dick Cheney’s life, including his business career with Halliburton, see the film *Vice* dir. Adam MacKay (2018).
disaster. In this case, much of the relevant capital stock had already been publicly provided (such as the parks and other physical land-and-capital stock of the borough), and of course the risk still lay squarely with the local authority, which ultimately stood as the guarantor of the public serviced that it had contracted-out to BRETS. Now Heath might tell me that all of this can be true here without it necessarily being the case that the need to pay out profits to private firms in public service provision will always lead to worse outcomes for service users, and in a strict sense Heath would be right. But I would at least want to suggest that a company like Halliburton does not get involved in local-level sanitation or parks maintenance out of a sense of public mission, but out of a vulpine concern to squeeze money wherever it can. And so while citizens may not be entitled to the ‘free lunch’ of knowing with certainty that public provision is always and everywhere better on the grounds of it not involving the extra cost of funding corporate profits, they are nevertheless entitled to a very strong suspicion against any way of arranging the provision of local public services that might contribute to the financial bonuses and lifestyle enhancement of someone like Dick Cheney.\textsuperscript{10} We should therefore be far more worried about the really-existing free lunches of corporate profiteers benefitting from public sector ‘derisking’ than we should be concerned about the conceptual ‘free lunch’ that Heath takes himself to discern in the blanket objection to the public funding of corporate profits.

VI. The Case for Caution on User Charges

We can now return to consider cases of ‘Type-1’ privatization in more detail, that is cases that involve “shifting from tax-financing to beneficiary-financing of state-provided goods” (Heath 2023, 41). Heath offers three rationales for when this kind of shift in financing can be justified: the first is to do with collective action problems, as when the introduction of road-pricing can help to reduce congestion; the second is concerned with avoiding forms of unwanted cross-subsidization, as with higher education user fees, if we are concerned that free or cheap university education ends up benefitting, say, graduates in medicine or economics at the expense both of those who end up in less remunerative careers, and those who do not attend university at all; the third reason, following the work of the economist János Kornai (1986), is to do with ‘soft budget

\textsuperscript{10} For more on the case for more democratic public services, and the case for insourcing see Guinan and O’Neill (2019a, 2019b); Labour Party Community Wealth Building Unit (CWBU) (2019), and Kishimoto et al. (2020)
constraints’, and the importance of bringing greater market rigor into public services that might otherwise endlessly increase in budget for little discernible benefit.

On user fees as a way of solving certain coordination problems, there can indeed be much value in the introduction of such fees in some cases. Road pricing is a good case, although this is much more convincing when it is done as one sees it practiced in Portugal, with an efficient automated system using numberplate recognition, as opposed to the less automated versions one often sees in the United States or France. When done badly, road pricing gives us long queues at tollbooths, that generate a huge destruction of economic value by slowing down the journeys of road users, and alongside this generate more avoidable misery through creating one of the dullest and most depressing jobs that exist in advanced economies: that of tollbooth attendant.

On the alleged harms of 'cross-subsidization' in cases such as higher education, Heath moves too quickly, and does not make his case convincingly. As I say above in section II it is a non sequitur to move from the (true) claim that higher education provides a distribution of both public and private benefits, to the questionable claim that the costs of higher education should therefore be divided in proportion to benefits received. A society that decides to make university education free could be expressing the value that it places on learning and self-development, rather than understanding what it is doing simply as reallocating earning potentials. Similarly, it may also be giving a special priority to ensuring that those disadvantaged potential students who would be most intimidated by the high 'sticker price' of a degree priced in proportion to its likely future private benefit would not be dissuaded from attending university.

Heath gives the example of his wife, whose medical degree at McGill University cost the same as Heath’s own philosophy degree, and expresses concern at “the multiple forms of unjustifiable cross-subsidization in this arrangement” (43, footnote 11). I am less worried about this ‘cross-subsidization’ for a number of reasons: firstly, one could instead simply think it was a wonderful mark of civilization that a country could train brilliant medical doctors without any significant financial barrier to joining the profession. Secondly one might plausibly think that the public benefit of training doctors was significantly greater even than the benefit of training political philosophy professors, and so there could be little objection to public funding for their professional formation. Thirdly, if one were worried about resulting distributive inequalities associated with
possession of a medical degree, this would always be something that
could be addressed downstream by the tax system, rather than something
that would need to be addressed by means of user charges for those re-
ceiving higher education. (Indeed, there is an interesting tension between
Heath’s readiness to embrace ‘upstream’ solutions regarding user fees in
higher education, rather than waiting for redistributive corrections, when
held against his favoring redistributive solutions rather than (upstream)
decommodification in other cases (on which see section III above). This
might suggest that there is altogether too much (insufficiently justified)
enthusiasm for the ‘commodity form’ in Heath’s approach).11 Fourth and
lastly, if one is worried about trained medical personnel who might then
leave the country after their subsidized public education, there are always
particular ways in which those problems could be addressed via minimum
public service requirements for certain professions (see Stanczyk 2012,
among others).

On Heath’s third kind of reason, the Kornai-esque concern with runa-
way public budgets, I would not want to demur entirely, but I would at
least want to register a note of caution. It seems axiomatic that we would
not want to promote waste in the provision of public goods and services,
but some apparent forms of ‘waste’ may have deeper rationales. Consider
a public mail system that operates on entirely commercial terms, without
any form of egalitarian regulation. Such a system would presumably make
it cheaper to send mail between major cities, while driving up the cost of
rural deliveries. That might be more ‘efficient’, but it would likely under-
mine the sense of status and standing of those in rural communities, and
communicate to them that they simply had no expectation of accessing
basic, essential services on the same kind of basis as others. This is likely
to be injurious to self-respect, and undermining of a sense of shared cit-
izenship. Similar arguments could be advanced for the idea of public ac-
cess to broadband. The British Labour Party’s 2019 manifesto included a
policy of public broadband provision, a policy which attracted much neg-
ative comment at the time, but which began to look rather prophetic in
light of the overwhelming significance of access to broadband during the
Covid pandemic, for education, social life, and public participation (La-
bour Party 2019). Relatedly, Heath mentions the case for closing under-
derused rail branch-lines, and the case for running commercially oriented

11 On upstream ‘predistributive’ approaches vs downstream redistributive approaches,
see O’Neill and Williamson (2012); Hacker, Jackson, and O’Neill (2013); and O’Neill
(2020a).
services as opposed to the dreaded ‘soft budgets’ of public rail companies. But there may be good egalitarian reasons to keep a rail system working, rather than judging each of its elements on ruthless commercial criteria, for reasons analogous to those supporting a national mail system. Moreover, a democratic society might also want to be able to take a long-run view on the preservation of certain public services which might need to be protected for potential future use and development, rather than thrown in the dustbin based on their performance in the current financial quarter. From the long-run point of view of the state and its citizens, it is much more costly to reopen or even to rebuild a branch line in the future, rather than to bear a few years of modest losses, whereas commercial organizations will find such long-run thinking alien when held up against the standards of the current financial bottom line.

With all three of Heath’s reasons in favor of expanding the role of the market in public provision (‘type-1 privatization’) he does indeed identify plausible reasons that ought to bear some weight in public deliberation. Nevertheless, Heath repeatedly moves too quickly from plausible pro tanto considerations to illicit conclusions, and is too slow to see the countervailing considerations that might lead us to regard some of his purportedly ‘anodyne’ privatizations as being significant policy mistakes. These mistakes are borne out of an unwarranted restriction on the range of values that we need to consider, a lack of attention to distinctively social egalitarian considerations, and a sense of being too relaxed about the normative and temporal distortions that can come along with an overly ‘commercial’ mindset.

VII. PUBLIC PROCUREMENT, INSOURCING AND OUTSOURCING

I have already given some reasons above, in Section V, for why we should be extremely cautious about cases of ‘contracting out’ public services, whether at the national level, or at the local level (as in my example of the unwelcome activities of Halliburton in my home borough in the 1990s). Making room for private profit will typically create costs that ultimately fall on service users and on those employed in those services. Moreover, as Joe Guinan and I (2019b) have argued elsewhere, it will typically reduce the degree of democratic control that a government—whether at the national, city or local levels—has on the way in which those services are provided and develop over time. Nevertheless, I am in significant agreement with Heath when he points to the ways in which existing government procurement rules are often at the root of significant problems in
public service provision, regarding relationships to suppliers, and that often the reason that more extreme forms of outsourcing become the only viable solution are due to the rigidity and perversity of existing procurement frameworks.

Heath rightly points to the way in which “government [...] is usually constrained by the requirement that it take the lowest bid on any tendered contract” (2023, 46), which militates against taking a broader or more nuanced approach to social value, and undermines the possibility of government agencies building up more constructive relationships with especially reliable and effective suppliers. As Heath points out, this often pushes government towards looking for extreme solutions (such as comprehensive outsourcing at the level of public functions) to solve these procurement problems that occur at the more micro-level. That said, it should be emphasized here that the same kind of ‘best unit value’ bottom-line thinking can just re-enact these same kinds of problems at the more macro-level (which brings us back to Ealing, BRETS, and Halliburton).

The solution here is not to encourage macro-level outsourcing, but instead to argue for a more sophisticated approach to public procurement. This is actually something that we already have seen in the United Kingdom, where the 2012 Social Value Act allows a more nuanced and holistic approach to public procurement, in which a broader set of considerations around job quality and other desiderata can be used in preference to just focusing on lowest unit price. This kind of institutional flexibility with procurement has allowed the development of more ambitious and progressive approaches to procurement policy, as part of the movement towards local ‘community wealth building’ in local and regional economic policy, which I and others have discussed at greater length elsewhere (Brown and O’Neill 2016; Howard and O’Neill 2018; Guinan and O’Neill 2019a, 2019b). On this issue, I am sure that Heath would not be in principle opposed to embracing this richer and more effective approach to public procurement, as opposed to treating reductive procurement rules as an element of bad policy which must be treated as fixed and then worked around (see also Davies 2013). A successful shift from outsourcing to ‘insourcing’, when alloyed to a more sophisticated framework for procurement, allows local authorities to exert more democratic control over service provision, and allows the pursuit of important
values of social equality and inclusion, not least in the promotion of higher quality, unionized jobs.\textsuperscript{12,13}

On public housing, I must admit that I am perplexed by Heath’s claim that there are “not really significant market failures in this sector” (2023, 50), by his downplaying of the degree of market power that landlords exercise over tenants, and by his characterization of the normative questions of housing as reducing to “essentially redistributive” questions (50). Given his characterization of what is at stake with housing, Heath feels comfortable in returning to the default ‘economistic’ position of looking to minimize on market regulation while dealing with any issues simply by adjusting the background distribution of purchasing power. It would take a long discussion to explore all the ways in which this picture is misleading, but suffice for now to say that on my view we need to acknowledge at least two normatively salient features of housing as a policy problem, both of which Heath seems to miss. Firstly, the social relation between landlords and tenants is often a socially inegalitarian, asymmetric relationship where one side exercises power and the other bears the consequences, and hence it is objectionable on social egalitarian grounds (McTernan et al. 2016). Secondly, our concern with housing as a social problem is not just about all-things-considered distributive outcomes, but is also about the way in which housing policy constructs the physical spaces in which people and communities live their lives, and in which they encounter one another. To take just one example, there is all the difference in the world between policies which lead to strict spatial segregation by class and race, and housing policies that bring those of different ages, racial groups, and social and occupational backgrounds into daily contact.

\textsuperscript{12} For an approach to ‘insourcing’ over ‘outsourcing’ that has been greatly influenced by work on community wealth building see Labour Party Community Wealth Building Unit (CWBU) 2019.

\textsuperscript{13} On the subject of unions, I do not have space to discuss at length Heath’s rather critical and disparaging account of the role of public sector unions, which he sees as often acting as a block to the delivery of high-quality public services. This was a familiar theme expounded by Tony Blair, and Heath’s account itself is clearly quite influenced by the work of Blair’s advisor Michael Barber (2015), who is quoted approvingly by Heath in a number of places. I am struck that Heath’s (and Blair’s) account here is rather one-sided, and that my own experience of public sector unions in the UK, for example in the health and education sectors, is that they are more often energetic defenders of what is most valuable in the public services in which they work and to which their members often display heroic commitment. On the role of unions in correcting for the disproportionate political influence of the interests of capital, see O’Neill and White (2018).
with one another. When the politicians of Red Vienna built the magnificent Karl-Marx-Hof housing development in the Geiligenstadt district, and built the other Gemeindebauten in their city (Gruber 1991; Blau 1999; Haderer 2023), they were not (pace Heath) making a mistake in illicitly intervening in a market that would be better left in the private domain. Neither is it any coincidence that ambitious egalitarian and social democratic governments have often put housing provision near the center of their programs (Broughton 2018; Hanley 2017; Hatherley 2020; Sennett 2018). Much, much more could be said on this issue, but my minimal claim is that any serious effort to create a more egalitarian society will involve significant public sector intervention in housing regarding its construction, provision, distribution, and regulation.

VIII. THE HAZARDS OF PARTIAL PRIVATIZATION, AND THE DEPRADATIONS OF PRIVATE FINANCE INITIATIVES

As I am a great admirer of Heath’s range and power of insight as a political philosopher, I am eager that my assessment of his view need not be excessively negative. So let me register here two further areas of strong agreement. Firstly, Heath is right that one particularly pernicious phenomenon is the kind of ‘partial privatization’ we often see when the quality of a public service is reduced to the point where people of means find private sector alternatives (private security, private medicine), which “leaves relatively powerless citizens consuming the public version of the good, people whose complaints can more easily be ignored” (2023, 52). Services that are only for the poor quickly become poor services. This insight, though, is perhaps more significant than Heath realizes, as it in general pushes us towards universal public provision, and a lack of ‘means-testing’, on both distributive grounds and on social egalitarian grounds. Heath should think of what would happen in a university that followed his injunction against ‘cross-subsidization’ with fees allocated according to financial benefit, where the mathematical economists would pay a fortune (at the level of their projected future marginal benefit), presumably be taught in shiny new buildings, while the students of social work or of world literature are presumably sent out to be taught in Portakabins.

14 Housing is a subject to which analytic political philosophers really need to give rather more attention; for two significant recent articles on this subject by Katy Wells see (2019, 2022). See also Bell and Zacka (2020).
Heath also gets things right (perhaps more than he realizes) on the depredations of ‘New Public Management’ theory, with its recalcitrant enthusiasm for bringing in the private sector even in cases where plausible commercial contracts could not be designed, and where private sector involvement would in fact simply be liable to result in egregious forms of value extraction. Perhaps the apogee of this nonsense came with aspects of Public Private Partnerships (PPP) and Private Finance Initiatives (PFI), where private sector value-extraction was often encouraged simply as a way of keeping down headline levels of public borrowing, even in cases where it was clear to all concerned that the purely public option would be significantly cheaper and more flexible in the long-run (Foot 2004). Here, as with some aspects of procurement rules, it seems that the most pressing need is simply for wiser and less ‘gameable’ forms of regulation, and for accounting rules that do not incentivize irrational transfers to extractive profiteers. 15 Given Heath’s clear-eyed appreciation of the self-defeating inefficiency of many of these Blair-era public management models, it is again surprising that he is not more hostile to private sector involvement in public provision than the position defended in his essay.

IX. TECHNOLOGICAL CHANGE, LIBRARIES, AND PUBLIC SPACES

In this final substantive section, let me turn to one of my great loves—indeed, one of the greatest kinds of institutions that humankind has ever constructed—public libraries. Heath is more skeptical about these institutions and their continued relevance, and thinks that their time may have come and gone, making use of an analogy with Montréal’s formerly magnificent public baths. In the early 20th century, Montréal’s public baths provided a wonderful facility to enhance the quality of life and social integration of the city’s poorer residents, but fell into abeyance with the development of higher standards of indoor plumbing. On libraries, says Heath:

While I love libraries almost as much as Martin O’Neill, I am not persuaded by his claim that they are public because “there is something transcendent, something deeply valuable, indeed almost miraculous, about a shared public space devoted to a role as a repository for

15 For some sense of the irrationalities of PFI in Britain, and its harmful and extractive consequences, see some of the many articles from The Guardian newspaper collected here: https://www.theguardian.com/politics/pfi. See also Monbiot (2000), Pollock (2004), McKibbin (2007a, 2007b).
human learning and memory” (O’Neill 2020b, 187). I can imagine a similar paean having been written 80 years ago about the solidarity-enhancing qualities of communal bathing, and yet no one would think, in the present day and age, that this constitutes a strong argument for public baths. (2023, 58, footnote 19)

On Heath’s view, the onward march of technological and social change will leave libraries as an irrelevance, similarly beached by history.

There is much that could be said in response to Heath’s argument-by-analogy. Firstly, despite the availability of e-books and PDFs, the book has proved a remarkably resilient piece of technology; it is not going anywhere, and for long as it does not, the case for libraries remains. Secondly, public libraries are not just about books. Let’s take two of the most impressive public buildings that have opened in Europe in recent years, the new central library in Helsinki, Oodi (Helsingin keskustakirjasto Oodi), which opened in December 2018; and in Oslo, the new Oslo central public library, the Deichman Bjørvika (Deichman Bjørvika bibliotek), which opened in June 2020. Both Oodi and the Deichman Bjørvika offer meeting spaces, studio and rehearsal spaces, 3-D printing, language cafes, parent-and-toddler story times, venues for readings and speaker events, and many other forms of engagement and interaction. Oodi tellingly describes itself as “a living meeting space”, “a non-commercial, urban public space that is open to all”; a description that gets to the heart of its value as an institution.16 It is a place in the city where people can go without their activity being focused on consumption and without being objects of potentially profit-making for commercial interests, but simply where they can go and engage with others, in a venue that is dedicated to ideas, social development, and creativity.

Democratic societies need those sorts of spaces, where people can meet each other not as consumers or competitors, but as fellow citizens. Some of that need can be met by this new generation of expansive, ambitious libraries, but in truth we need much more collective public space in our cities: community centers, youth clubs, museums and galleries, cultural centers, parks and city squares, and all the other places in which we

---

16 For the Oodi (Helsinki) and its self-description, see https://oodihelsinki.fi/en/what-is-oodi/. As the website goes on to elaborate, “Oodi provides its users with knowledge, new skills and stories, and is an easy place to access for learning, story immersion, work and relaxation. It is a library of a new era, a living and functional meeting place open for all”. For Oslo’s Deichman Bjørvika, see https://deichman.no/.
can come together as citizens rather than market participants.\textsuperscript{17} And so, thirdly, like Heath I too can imagine an 80-year-old paean to “the solidarity-enhancing qualities of communal bathing” which may, in the era of indoor plumbing, feel as if it would have been beached by history; but such an argument would remain more timely than Heath allows (58, footnote 19). On the one hand (keeping the link to immersion in water) public and municipal swimming pools and leisure centers remain an important kind of institution, not least when they spread opportunities for certain kinds of leisure to those who might not otherwise be able to access them, and when they bring together people of different classes and backgrounds. In 1941 George Orwell could write in \textit{The Lion and the Unicorn: Socialism and the English Genius} of “the naked democracy of the swimming-pools” as precisely the kind of phenomenon that worked to break down class divisions between citizens, and to create a society in which “the same kind of life […] is being lived at different [income] levels” (33). On the other hand, even when the \textit{precise} form of shared public spaces may change over time, in part due to the ongoing processes of social and technological change that Heath describes, the need for \textit{some kinds} of such places and such shared institutions resiliently remains.

\textbf{X. CONCLUSION: REASONS TO REJECT PRIVATIZATION IN DEMOCRATIC SOCIETIES}

I am conscious that my aim in this discussion has often been to highlight differences rather than commonalities between my view and the position defended by Joseph Heath. But there is much on which we agree. Heath himself emphasizes that he does not seek “to defend privatization as a panacea”, and he allows that many real-world privatizations have been “terrible, ill-considered or even flagrantly corrupt” (2023, 59). On those points we agree, and I can even agree in addition that at least a subset of his purportedly ‘anodyne’ privatizations are not in fact objectionable (think, for example, of Portuguese road pricing). And as I said at the outset, Heath’s provision of a ‘grid’ or conceptual map of the territory has been a valuable intellectual public service.

But nevertheless, these finding themselves carried along by Heath’s more reasonable moments should nevertheless be prepared to dig their heels in where necessary. Heath’s economistic framework for thinking about these issues rules out too many important normative

\textsuperscript{17} For a convincing case in defence of the importance of physical public space even in the digital age, see Parkinson (2012).
considerations from the outset, and biases the discussion away from considering the full diversity of reasons against privatization in its different forms. Heath writes that he would have progressives accept the “provocatively ambiguous German Social Democratic Party slogan: ‘As much market as possible, as much state as necessary’” (59; Crouch 2021). On my view, something closer to the obverse of this tepid slogan would get us much closer to a defensible normative position: ‘As much public provision as possible, as much market as necessary’. Markets are tools to be deployed by democratic publics and their governments when they have good reason to use them; they should not be a kind of normative default. There is much more that we can and should demand from our public institutions—in terms of creating the preconditions for democratic deliberation; in terms of generating social solidarity and inclusion; and in allowing citizens to relate as equals—that is not captured by the Heathian view that sees public provision as always and only a response to specific kinds of market failures. As free and equal citizens in democratic societies, we are more than mere market participants, and our institutions should be constructed in a way that treats this fundamental truth with due seriousness.

REFERENCES

Heath is right that such a slogan is “provocatively ambiguous”, of course: in what way “possible”, and as “necessary” for what (59)? One is reminded of Tony Blair’s purportedly anti-ideological slogan that “what counts is what works”, but what counts as what works in the relevant sense, and to what end? As the title of an essay by Ross McKibbin (2008) put it, “what works doesn’t work”.

18


Boston Review. 2 June 2016.


**Martin O’Neill** is Professor of Political Philosophy in the Department of Philosophy at the University of York.

Contact e-mail: <martin.oneill@york.ac.uk>
Ethics from the Outside Looking In: An Interview with Roger Crisp

ROGER STEPHEN CRISP (Brentwood, 1961) is Professor of Moral Philosophy in the Faculty of Philosophy, University of Oxford. He is the Director of the Oxford Uehiro Centre for Practical Ethics and a Tutorial Fellow of St. Anne’s College, Oxford. He holds an Honorary Professorship at the Dia- noia Institute of Philosophy, Australian Catholic University. Crisp received an MA in classics from St Anne’s College, Oxford, where he later took both a B.Phil and D.Phil in Philosophy. For the B.Phil., he was supervised by J.L. Ackrill, Jonathan Glover, James Griffin, and Alan Ryan. Crisp’s doctoral thesis (1988) was on “Ideal Utilitarianism” and supervised by James Griffin, Michael Lockwood, Derek Parfit, Joseph Raz, and David Wiggins.


The Erasmus Journal for Philosophy of Economics (EJPE) interviewed Crisp about his formative years (§1), his work on utilitarianism and the
virtues (§II), the good (§III), practical reason (§IV), future research projects (§V), and his advice for graduate students (§VI).

I. Formative Years

_EJPE:_ Professor Crisp, what first sparked your interest in Philosophy?

ROGER CRISP: I don’t think it was my parents, though we did discuss political issues. I went to a school with many outstanding teachers. The best, and the one who most influenced me, was a great classicist, Dennis Rid-diford. Dennis did tell us about Plato, and I found that interesting. People often say that children are naturally philosophical, and I think that I had already started thinking in an abstract way. Then I went to university. I came to Oxford to study classics. It’s a tripartite degree, involving literature, history, and philosophy. I had worked on literature and history, but I hadn’t done any philosophy. The first philosophy I studied was Plato’s _Phaedo,_ which is a very hard dialogue: I still find it tricky, so that was a struggle. The next thing I read was Locke’s _Essay._ I can remember quite vividly sitting in my garden for weeks over the summer wondering what an earth was going on in this book. It seemed to me like armchair psychology and that it was all just made up, because I was approaching it in completely the wrong way. By the end of that summer, I had decided that I was going to give up philosophy. I went to find my tutor, Gwynneth Mat-thews, who luckily wasn’t in, and my tutorial partner recommended that I give it another term. Then we looked at Mill’s _On Liberty_ and that’s when I decided I really wanted to do more philosophy. I think this was because Mill was discussing some of the things that I had been thinking about. Obviously he’d got a bit further(!), but that’s what excited me, especially the issues of free speech, duties to the individual, and the nature of the good society.

What views helped shape and influence your early thought?

When I was an undergraduate, I was required to write one or more essays each week on a separate topic, and I thought of it in a kind of serial way. My job each week was to come up with decent essays. Because we were covering such vast areas, and there was so much to do, I don’t think I ever did much global thinking. So what I was really trying to do was come up with a clear view on each topic, I suppose in the hope that everything would fit together. But to be honest, even that didn’t matter very much because all I was trying to do was come up with views I could use in exams, and what was important there was to make sure all the three essays in the exam were consistent with one another.

Then I did the BPhil. I remember the letter of acceptance from John Ackrill. My mother brought it to me, and it just said, “Dear Crisp, we have decided to admit you to the BPhil. Yours sincerely, John Ackrill.” (My mother explained that Ackrill’s addressing me by my surname was in fact rather friendly.) I was lucky enough to have Ackrill as my main supervisor.
He was wonderful. With him, I worked on Aristotle. And then I had Jonathan Glover for my thesis, which was on the minds of animals: mainly philosophy of mind and language. I had Alan Ryan for politics. The person who had the greatest effect on my thinking was James Griffin. Partly because he was Jim Griffin, but also because of the things he was working on, which happened to be what I was most interested in. And then, of the faculty members around at that time, R.M. Hare and Derek Parfit had the biggest influence on me. Partly through their writings, but also because they were very ready to talk. I should also mention how much I learned from many other graduate students. One of them was Brad Hooker and he has been an inspiration to me throughout my career.

Let me say one other thing about the B.Phil. I've talked about the topics I did and the subjects I covered with Griffin, but that of course leaves out metaethics. Griffin had a metaethical view. But it was slightly negative. He tended to think that many of the distinctions people use are unclear and that we could probably do without them. But I wasn't persuaded by that. I was interested in the debate between people like McDowell and Wiggins on the one hand, and Blackburn on the other. I was particularly attracted by a view—which both McDowell and Wiggins argued for—according to which evaluative properties, particularly moral properties, are importantly analogous to secondary qualities. For example, there's a paper by Wiggins called “Sensible Subjectivism” (1987), and also McDowell's “Value and Secondary Qualities” in the volume Honderich (1985) edited in honour of Mackie. My thesis, which was a bit of a grab bag, contained a chapter defending that view. But that's one of the positions I came to reject, I guess because, reading people like Clarke, Sidgwick, and Parfit, I came to the view that that kind of moral metaphysics is unnecessary, and we can avoid it by recognising that the reasons we have are capturable in statements of necessary truths. (I might add that I think Clarke and Sidgwick saw that more clearly than Parfit, who spent too little time thinking about the epistemology of normativity.)

**So how did you end up writing a thesis on Ideal Utilitarianism?**

I had enjoyed working with Griffin and knew I wanted to do moral philosophy. I don't think we were asked, “what would you like to do your thesis on?” You just said, “I would like to write a doctorate” and they would say, “Yeah, Okay.” So I started in the first year writing some stuff that was influenced by Dworkin's views of jurisprudence, trying to argue for a conception of morality where there were good and bad interpretations of it, and what we are looking for is the best interpretation of morality. But (and I'm not surprised looking back) that ran into the ground. But it helped me learn more about philosophy and how to do it, and Griffin was encouraging about it. He told me I'd made some progress and developed as a philosopher, so now I had to decide what to do next. Griffin was sharing a draft of his book *Well-being* (1988) with the large group of excellent graduate students at that time working on moral and political philosophy, and this got me interested in well-being. I was pretty influenced
by his criticisms of hedonism and desire-satisfaction accounts. Griffin sometimes says in the book that he accepts a version of informed-desire theory, but I don’t think he does. I think it’s an objective list account, but Griffin didn’t like that name, partly because he thought the distinction between value and desire breaks down. So that led me to try to defend an objective list theory. I was attracted to consequentialism (or utilitarianism) and Tom Hurka had heard I was interested in an objective list account, and suggested to me that I should read Hastings Rashdall’s _A Theory of Good and Evil_, which I did, and that led me to think, “Oh yeah, this all makes sense to me.” Well, not everything Rashdall said made sense, because he also tries to plug a pretty non-consequentialist account of virtue into his account of well-being. So I didn’t do that. I just defended this objective list theory and attached it to the utilitarian principle. And because I was interested in applied ethics, I included some chapters in practical ethics, one on vegetarianism, and the other on population ethics.

_I will ask you more about the upshot of your thesis shortly, but beforehand, looking back, what would you consider to be your early career breakthroughs?_  
I’m not sure it counts as a breakthrough, but looking back I would say that probably the Rashdall element was the most important. Not only the objective list element, but also his arguments for impartiality, in particular the analogies between time and space, which Parfit then developed as well. I think that was what stayed with me most.

_Which of your early convictions have you retained and which ones have you now jettisoned?_  
Well, I was fairly convinced by standard act utilitarianism in a way that I’m not now. But I still think that anybody ought to accept the utilitarian principle, as one principle among others. That is, if the principle is, other things being equal, produce more happiness than less, that seems to be pretty undeniable. But I’m inclined to think that utilitarianism doesn’t give enough weight to the separateness of persons. And this is one of the things I learnt from Sidgwick. I also gave up on the objective list account. I don’t think I realised this at the time, but I was on the one hand learning from Griffin, while on the other hand rejecting some of his conclusions, because, unlike many philosophers in ethics, he was prepared to take morality as a system—a bit like religion—and think about it independently. In other words, it could all turn out to be something we have to understand primarily from an anthropological point of view. This line of thought is also at least present in the work of John Skorupski, from whom I’ve also learned a great deal over the years. Indeed, it’s there in Mill, on whom of course Skorupski wrote, and in many other British moral philosophers.

_Do you mean that Griffin was prepared to take morality as a natural kind?_
I don’t think he would have put it quite like that. But very roughly, that is what I got from his approach. And, though he was not a utilitarian, he was ready to take it seriously in a way that many people weren’t. Utilitarianism is obviously a very radical position, because it says: “we’ve got this set of principles, which we call morality, but only one of them is correct, and you’ve got to take that out, and let everything else rest on it.” All of the other principles then become secondary. That’s essentially the two-level utilitarianism of the kind you get throughout the tradition. And what I now think is that that radical aspect of utilitarianism was carried across to the theory of well-being by Bentham, Mill and Sidgwick, in particular. They were ready to take the common-sense views about what matters in life—like accomplishing something, friendship, and so on—and take seriously the thought that they were mistaken, in the same way that common-sense morality is mistaken. There is one exception: when it comes to the hedonistic principle, we can’t give that up: pleasure is good and pain is bad. This is what led me in the direction of hedonism. So I think there’s a tension in the kind of view that I defended in my thesis, where you’re radical about ethics—you’re ready to be minimal and impartial—but when it comes to the view of well-being, you’re not. You’re actually just somewhat unreflectively adopting the view of common sense.

**Why, exactly, do you think there’s a tension there?**

Well, there’s a tension in approach. I’m not saying that you shouldn’t end up with an impartial, maximising perfectionism. But, if common sense has got it wrong about morality, it’s quite likely to have got it wrong about well-being also.

**We will pick up on this more, in a moment, but first, let me ask, how would you describe the philosophical environment at Oxford during the time of your first appointment?**

Very, very exciting. I’ve actually got all my notes somewhere in this room. I kept all my lecture notes, and one day I want to go back and look at them, because I’m sure I’ve forgotten a lot.

You may have heard of the seminars that came to be known as ‘Star Wars’, involving people like Parfit, Cohen, Sen, and Dworkin. The debates between McDowell and Wiggins, on the one hand, and Blackburn, on the other, were also gripping. Hare was around, Williams was around, and many others too. It was just a great period for moral philosophy, and I count myself extremely lucky to have been in Oxford during that period.

**So this was the late 80s going into the early 90s?**

Yes, that’s right. I had a British Academy postdoc at University College from ‘89, and then I started the job here [St. Annes] in ’91.

**In David Edmonds’ (2023) recent biography of Derek Parfit, Edmonds describes a regular gathering with yourself, Brad Hooker, Ruth Chang,**
and Derek Parfit. What were these meetings like? What (if anything) did you (dis)agree about the most?

It’s quite funny really because the person who had the idea was Ruth Chang. She had this nice apartment (or nice room) and so the first meeting was in there, and she brewed this great coffee, and everything went very well. Then Ruth said to Derek, “Shall we have another one?” And Derek said, “Yes, we should definitely have some more, but you can’t come.”

Parfit was Ruth's supervisor, and he thought that she was taking rather too long over her thesis (she might have agreed with that, I guess). So I think there was at least one meeting where she set up the coffee in her room, and then disappeared—which is a great pity, because it would have been great to have had her there! (Maybe she did come to a few later on.)

The people who were there included Brad Hooker, Philip Stratton-Lake, Jonathan Dancy, John Broome, Julian Savulescu, Ralph Wedgwood, and others. And then there were visitors as well, often people staying in All Souls, like John Skorupski, Tim Scanlon, Michael Bratman, Bob Adams, and others. The meetings tended to revolve around work-in-progress. We didn’t really read published articles. And it was somewhat dominated by Parfit. Probably half to two-thirds of the discussions were about his drafts. And, I think, for us, it was probably more stimulating earlier on, because Parfit started going around in circles a bit towards the end. But in the earlier sessions he was really working in linear ways, so that was very exciting and inspiring. He really was a phenomenon: much quicker than everyone else.

Were there any points of convergence? Which aspects did you really disagree about?

Hmm. I don’t remember a great deal of convergence. In fact I don’t recall a single occasion where we ended thinking, “Right, well we’ve got that sorted out!”

As I mentioned, it did revolve around the themes that people were working on. I remember that the notion of reasons was pretty dominant: I can recall one early session where I’d written something that came up with (I think) five different senses of reason, and Derek was very excited about that—probably with little justification. Still, people really hadn’t thought all that hard about reasons, though obviously there’s been a lot of helpful work since then. (I was very pleased to hear from Derek years later that I was the person who had first persuaded him that neutral subjective or psychological states such as desires can’t ground reasons for action.)

The meetings went on and probably became more spaced out. But we kept going, and then, I’m not sure when, but after quite a long time, Derek told us that he needed to stop because he wanted to finish his book. We were disappointed, partly because we thought he wouldn’t finish the book. But then of course he did—so maybe it was the right thing in the end!
II. UTILITARIANISM AND THE VIRTUES

As mentioned, you wrote your D.Phil thesis on ‘Ideal Utilitarianism’. What was the main upshot of that thesis?
The main upshot for me was that it got me thinking more about impartiality, and well-being. But of course in writing it, I read a huge amount and talked a great deal to other philosophers which really helped the development of my philosophical views in general.

In work leading on from that thesis (1992), you argue for a form of ‘Two-level Utilitarianism.’ Before I ask my next question, perhaps you could explain what you mean by two-level utilitarianism, exactly?
I don’t know where the ‘two-level’ phrase comes from. I think it comes from Hare, doesn’t it? And then Griffin took it on as well. Anyway, Hare got it right, really. The idea is that one level of thinking about utilitarianism—or ethics in general—is at the philosophical level, trying to get to the truth about ethics philosophically. And that concerns what people call—and this isn’t a term from Hare—the criterion of rightness. And, though I wouldn’t put it in terms of ‘rightness’ (because I think that reasons are fundamental), what you’d be looking for there is the truth about the reasons we have, i.e., what makes a particular action the one you have reason to do.

So that would be at the critical or foundational level, as it were, questions about what grounds one’s reasons for action. But what about at the second level, what goes on there?
The other level would be stuff that goes on outside the study, essentially. The distinction is a bit rough and ready, because of course you will in the study think about how you ought to act outside, and vice versa. But the crucial idea, as far as utilitarianism is concerned, is that when you leave the study, if you’re a utilitarian, it doesn’t follow that you should go around all the time trying to put utilitarianism into practice, because that may itself be something that utilitarianism recommends you shouldn’t do. This was Hare’s ‘intuitive level’. The philosophical level, he called the ‘critical level’.

Like most thinkers in the utilitarian tradition, Hare thought that common-sense morality—the morality that we’re brought up with—has quite a lot going for it. I think one failure in the utilitarian tradition (though there are exceptions to this, perhaps in the works of people like Mill or Singer) is that utilitarians haven’t said enough about how damaging common-sense morality is in various ways. In particular, common sense perpetuates the myth expressed in the so-called ‘acts and omissions’ distinction. It also tends to encourage us to focus on what we’re doing as individuals. So it leaves out collective aspects which are hugely important.
In the afore-cited paper, you argue for a position which you call ‘Biographical Utilitarianism’ (BU). BU holds that “Any individual ought to live in such a way that the total amount of utility in the history of the world is brought as close to a maximum.” (p. 141) You then claim that BU is consistent with, and most likely to recommend, a life of the virtues. But why do you believe that BU would end up recommending a life of virtue, in particular?

That paper (if I remember) was a response to some criticisms about various forms of utilitarianism—including act-, rule-, motive-, and other forms of utilitarianism—and I thought, well, what you could do is just bundle all of these things together into the context of a life, and then what you should be thinking about is leading the life that results in the best history of the world. I had some conversations a few years afterwards with Parfit, Michael Smith, Phillip Pettit, and some other people about these issues and what should be the focus of the utilitarian ethics. The upshot was that I came to believe that there is something special about acts, because acts very often result from decisions, and decisions are meant to be based on reasons. So actually, the question, “Which act should I now perform?”, is a pretty salient one. Hence the focus on acts in the ethical tradition makes a lot of sense. But you’ve got to be clear that the consequences of your act might include all sorts of things like what sorts of dispositions you might be consolidating or weakening, or what motives you might be making yourself more likely to have in the future, what your character is going to be like, and that’s where virtue comes in.

To answer your question, the reason I brought the virtues in was that it just looked plausible to me. I think that somebody with the virtues, properly understood, would have to be someone who, for example, recognised the myth of the act/omissions distinction—it wouldn’t be the typical virtuous person as currently conceived—and then I believe that person would be living the right life. They would have to be thinking about decisions that they are taking at particular times, which will result in their living a life of that kind, but their life would not be dominated by constant reflection on that life.

Thinking out loud, here—and let me know if you agree—but one way to cash out the idea would be to say that the virtues, or certain dispositions or motives to act, are thus prior to the acts themselves. Acts are important, but these stem from the dispositions, motives, or virtues that we might have. And hence why it is important to cultivate these dispositions, motives, virtues is because they help to select the right action. Exactly. But then you need to take into account the consequences of all of those. And I think that if you don’t get that distinction clear, you can end up muddled. So, for example, I think that Parfit’s (1986, chapter 1) idea of blameless-wrongdoing is a muddle because he didn’t get clear on the distinction between actions and dispositions. He was thinking about act utilitarianism, and the act utilitarian can take into account the effects of
actions on your dispositions in the future. So if you fail to benefit some stranger and instead benefit your child, you can take into account the effect of that action upon your future dispositions. Act utilitarianism has no problem with this. Of course, act utilitarians—as far as I'm concerned—shouldn't be talking about wrongdoing, they should be talking about what we have reason to do. And they certainly shouldn't be talking about blame. Blame is at the heart of common-sense morality, but it's really just a matter of anger and other emotional reactions. And when you should express anger is, again, a question that the act utilitarian can ask. You're not going to end up with anything like blameless-wrongdoing.

But would it not be more likely that BU recommends something akin to a non-principle-based deontology, such as moral particularism? That is to say, on a two-level view, BU would be true at the 'critical level', but would recommend that an agent ought to be responsive to particular moral reasons or considerations at the 'intuitive level'.

To frame the question slightly differently, one might ask, isn't positing the virtues unnecessary here when all that seems to be required is good moral judgement? And if not, why not?

Another good question. It depends on what you mean by particularism. I think when Dancy is talking about particularism, it's a kind of meta-view. It's not a meta-ethical view in the sense in which most people understand that term—it's not a metaphysical view—but a view about moral principles. It says that the truth about ethics is not captured entirely by moral principles. So I presume that BU (or any of these views) would not be consistent with that.

But if we go back to the distinction between levels again, then I think you are right about that. In other words, BU (or any other plausible form of utilitarian principle) wouldn't be recommended. Rather, the implication would be that what you ought to do is just decide what to do in any of the particular cases, with some degree of appropriate reflection.

Yes, this is what I had in mind when I suggested good practical judgement. Notwithstanding, implicit in my previous question, I guess, is a desire to get your thoughts on what you think a virtue actually is. Because, at least on my reading, it can often seem as if Aristotle wants to equate virtue with good practical judgement. If this were so, and if we were to accept this picture, then my last question would not be so much of a criticism of your view, but rather another way of phrasing it.

I think, roughly, Aristotle got it right on the nature and structure of virtue. He's the greatest theorist of the virtues. And it's sad that, despite the revival of virtue ethics, many aspects of his view have been misinterpreted.

The distinction between actions and feelings is clearly very helpful. And it's illuminating to think about life as consisting in certain spheres, like money—your control of money—and the emotions, anger, fear, and so on, in each of which what you should be aiming at is getting it right. You can get it wrong by failing to do the right thing at the right time, and
so on, which is Aristotle's set of defective vices. And you can do the act in question at the wrong time and the wrong place, and so on. So, if we take generosity, the generous person is the one who gives away the right amount of money, at the right time, for the right reasons, and so on. The prodigal person is the person who does it when they shouldn't, and then the stingy person is the person who fails to give away money when they should. Aristotle recognises in his discussion of generosity that it's going to be quite common for a person to have both vices. So it's not as if you can place each person on a spectrum. If you're prodigal, then you'll give away all your money. Then someone will come around collecting for a worthy charity, and you can't donate to them, even though you'd like to, because you've given away all your money, and that makes you mean or stingy. Aristotle doesn't have the focus we have on the will and intention. It's your fault that you cannot now donate. Thus you've got both vices. I don't know what he would say about someone who had both vices and gave away money at the right time. I presume he would say that they're not even a little bit virtuous. Virtuous people don't just get it right, they also don't get it wrong.

But I guess my question is that people often tend to list the virtues—generosity, courage, bravery, and so on—but these, at least to my mind, seem to be instances of what is just one virtue, namely, the virtue of practical wisdom, or, more loosely, the virtue of getting it right.

Absolutely, yes. So Aristotle's brief account of moral education is an account of learning to get it right. People become virtuous by being properly taught, either deliberately or by inhabiting a particular culture, and so they're naturally disposed, say, to give away money. But that's not enough because you cannot just let someone off if they're naturally disposed to give away money since they might give it to people who shouldn't have it. And presumably they won't give money to everybody, so they might fail to give money to someone they should. That's why they have to develop what Aristotle calls the intellectual virtues. Next, though, and this is where I disagree with Aristotle a bit, he follows through with the Platonic/Socratic view, and says that, if you're going to have one virtue, then you must have all of them. And you can see why he thinks that, because, if you don't have all of them, you're not insured against cases where you might not do the right thing. You might be thinking that you're pretty generous because you give away the right proportion of your income in certain circumstances. However, to do this consistently might require a certain amount of courage. And if you haven't got that virtue, then you actually fail in the sphere of generosity. So on Aristotle's view you haven't got the virtue of generosity, since it's success-oriented. And I can see the argument for that. But the reason I disagree with Aristotle is because he's just a deontologist. He thinks that there is a right way to act, and a right way to feel, and so on, for any particular occasion. And given this, he comes up with an account which is pluralistic in terms of the set of principles corresponding to each virtue. Whereas I'd be more inclined to think
that the correct principle is consequentialist, really. Ethics is essentially
to do with the distribution of well-being. So the question of whether you
should have all the virtues, or just some subset depends on what the out-
come is of having those virtues.

In your paper “Utilitarianism and the Virtues”, you state that “an agent
ought to live virtuously, consulting the BU criterion only on certain spe-
cial occasions.” (p. 154) But I wonder why this latter qualification is
necessary. To be sure, there may be times when the right thing to do is
to straightforwardly maximise the good. But surely a virtuous person
will already be disposed to act in such a way without having to “consult
the BU criterion.” So why do you believe this latter qualification is
needed?

Well, it might not be. But I take it that people like Mill, and Hare are right,
in that it would be probably a mistake to think to yourself fairly early on
in your ethical career, “Well, this common-sense morality that we’ve got,
it needs a bit of tweaking, but I can now tweak it and live in accordance
with that for the rest of my life”. That seems a bit complacent. The mo-
rality you’ve been taught may be appropriate to the time in which you are
living. But things may change. Or, of course, you may have got it wrong.
So, I think there is monitoring that has to be done, and there might be
conflicts that require some kind of resolution.

And do you not worry that the recommendation to live a life of virtues
whilst consulting the BU criterion on special occasions might lead to a
form of cognitive dissonance on behalf of the agent?

Yes, it could lead to such an incoherence, and that would be another fac-
tor I suppose for someone to take into account when questioning how
one ought to live. In fact, I remember talking to Hare about this—well,
several of us talking to Hare about this. He and his wife, Catherine, really
loved one another—they had a wonderful partnership—and he was pre-
pared to be very frank, so we said to him, “Isn’t there some tension be-
tween the way you feel about Catherine, and your thoughts about ethical
theory’s being completely impartial?” And without hesitation he said,
“No”. I take it that he really was ready to draw a sharp distinction between
the study and the rest of his life. It wasn’t as if he was living his life and
then he went back into the study to think about utilitarianism. He
wouldn’t be thinking to himself that in some sense he oughtn’t to have
been doing those ordinary things in life; he was convinced that he lived
in accordance with his theory, at least in that respect, but probably in
many other respects.

How important do you think it is for philosophers—especially practical
philosophers—to live in accordance with their own philosophical views?
One example that comes to mind is G. A. Cohen (2000), in his well-
known paper, “If you’re an egalitarian how come you’re so rich?” An-
other example perhaps being some more hard-nosed utilitarians who
openly believe that they should be acting more morally in their own lives, but just don’t. Do you have views on this? I don’t myself think there’s anything bad in itself about such inconsistent or hypocritical behaviour. Consider some appalling fascist, who acts in accordance with their moral views. I wouldn’t want to say, “Well, of course what they’re doing is overall wrong and regrettable, but at least they have the virtue of integrity”. We also have to be careful about what the implications of a theory actually are. A very rich egalitarian might be developing some fund which will promote equality much more effectively than their handing over their money piecemeal to one charity or another. What people have reason to do is what they have reason to do, and this may be in line with their own view, or it may not. But its being in line with their own view doesn’t give them an extra reason to do it.

Moving on to virtue ethics more generally, you mention in the article that the value of the virtues is derivative from the value of welfare. In a different paper, “A Third Method of Ethics?”, you argue that virtue ethics as a theory of right action should be categorised as a ‘non-principle-based deontology.’ However, you entertain the idea that virtue ethics constitutes its own branch (or method) of ethics if considered, not as a theory of right action, but as a theory about what type of characters we ought to possess. You reject the idea that there is added moral value in a particular act simply because it was performed from a virtuous motive (or disposition). Some philosophers, though, hold that being virtuous is intrinsically or non-derivatively valuable (for example, Hurka 1992). I must admit, I feel the force of this view. Maybe you could say more about how you see the relation between virtue (or the virtuous person) and welfare and why you think that being virtuous is not ultimately valuable?

In the 80s, I edited a book called How Should One Live?, a collection on the virtues. I haven’t looked recently at the introduction to that, but I suspect I wouldn’t agree with all of it now. One of the things it says is that this turn in thinking about the virtues has been hugely productive in moral philosophy. I haven’t given up on that. I think the issues raised about the relation between the self and others, about partiality and impartiality, the nature of dispositions, the two-level issue, how to understand the virtues, and which virtues we now live by, and so on are all worth thinking about.

---

1The argument, roughly, is that, when pressed to explain what makes an act right, the proponent of virtue ethics is led to the view that a virtuous act is so by dint of its being right. For example, it would be virtuous to save a drowning child because, under certain conditions, this would be the right thing to do. Pressed on what makes it the ‘right thing to do’, the virtue ethicist would then draw attention to various normative factors which she deems relevant. This, therefore, takes the form of a non-principle-based deontology. And so, as a theory of right action, virtue ethics is not a distinct method of ethics alongside consequentialism and deontology.
The turn in virtue ethics that I disagree with is the idea that there is a third kind of ethical theory which nobody had noticed, though Aristotle got part of the way there. That strikes me as a mistake. Because, once you understand what ethics is aiming at, you see that it’s trying to state clearly and precisely what your reasons are. If you’re a virtue ethicist, you have to say that a virtuous person’s actions are right because they’re virtuous. And that’s terribly circular and implausible. It also gives you a weird account of the virtuous person because in most cases, if you ask a virtuous person why they are acting as they are, they won’t say, “Well, I’m doing it because I’m virtuous” or “I’m doing it because it is what a virtuous person would do in these circumstances”. They’ll say, “Well, I’m doing this because it’s just” or, “I’m doing this because it would have been unkind not to do it”, and they can articulate without reference to virtue itself what makes the action in question virtuous. So, the virtue of kindness just consists in doing kind actions. And that’s really my objection to so-called ‘virtue ethics’. That seems such an obvious problem with virtue ethics that I’d be surprised if it continues for much longer. It’s pretty well entrenched at present, but so was phrenology.

One thing I was also trying to do in that paper is to say that it’s really worth thinking about the virtues because they’re the property of all moral philosophers, not just those defending one particular theory. And it certainly is worth asking the question which the ancients asked, namely, “Should we say that all that matters is doing the right thing, or perhaps having the right feelings on individual occasions, or should we say there is some value in having the disposition to do that?”

I remember years ago having a conversation with Frances Kamm, who has probably written more about deontology in the last few decades than any other well-known philosopher, and I asked her why she never talks about virtue. And she said, “Well, why should I? If somebody is living their life in such a way that they always do the right thing, for the right reasons, then, what else is there to say?” I can understand where she’s coming from. But I also think the questions that the ancients asked, and that W. D. Ross asked, such as “Is there value in having virtuous dispositions?”, is an important and interesting one. Ross thought, yes, there is value in having these dispositions. So he was actually a virtue ethicist in that sense.

What do you mean by ‘value’ here?
I mean that it makes the world better—that it’s a good-making feature of a person.

And, if I’ve read you correctly, you would agree there is moral value here, but not non-ultimate value, as you put it.
One view I’ve come to, mainly through reading Sidgwick, is what’s come to be called the demoralizing position. So, if you’re going to capture reasons for action, it’s best to try to avoid using moral terms. This is partly for reasons of parsimony. But there are other reasons that we might talk
about later. Then you’d have to ask the question without reference to moral terms. And I think it starts sounding less plausible. Also, I’m a welfarist, so I don’t believe in the idea of any kind of moral value that is not to be explained by reference to well-being. I grant that such value is conceivable, but I guess that welfarism is more aligned with the theory of reasons.

Thus, in your view, to be a virtuous person has value only to the extent that it increases my well-being?
And the well-being of others, of course.

III. THE GOOD

This brings us onto the topic of welfare or well-being more generally. You are well-known for your writings on Hedonism. What do you find attractive about this view, and why do you favour it over other leading theories of welfare (such as ‘desire-satisfaction’ or ‘objective list’ theories)?
The way I see the history of discussions about well-being is that the debate (at least in the Western tradition) has been between hedonism—which was hugely popular in the ancient world—and objective list theories. The objective list theory was defended by the three philosophers who dominate discussion of ancient philosophy, Socrates, Plato, and Aristotle. And I think that debate between hedonism and the objective list theory runs through Christianity. There’s a kind of side debate about whether we include virtue on the list. And I think that when we get to the so-called British Moralists, the majority of them were hedonists. But hedonism started to become less popular in the 19th century. Why that happened is a bit of a mystery. Terry Irwin suggested to me—in what was just a throw-away suggestion of his—that it might have had something to do with Victorian sensibilities: it was not thought of as being sufficiently high-minded. And obviously there were the worries about sensualism and so on in the background. And then, of course, we have Moore—who was not a hedonist—and he really put the boot into Mill in a not terribly plausible way, but in a way which was hugely influential. Still, I’m inclined to think that hedonism is on the rise again. The truth will out!

As I mentioned, I defended an objective list theory in my thesis. But then I moved away from it as a result of thinking about the common-sense view of well-being in the same way that I thought about the common-sense view about morality.

Most objective list theorists have said, and would say—though there have been exceptions, not least, Aristotle—that pleasure is good and pain is bad, other things being equal. Thus there is agreement there. And then the question is, should we say that there is other stuff that’s good—accomplishment, friendship, knowledge, and so on?

---

2 See, for instance, Crisp, (2006a, Ch. 6; 2006b).
One thing that happened towards the end of the 19th century was that people started to argue that we shouldn't be thinking in either of these ways, but that we should be thinking in terms of desire- or preference-satisfaction. That came out of philosophical economics. And the view became very popular in philosophy in the last century, and is still quite widespread, though I'd say it’s losing ground in philosophy. That view seems to me seriously mistaken. It misunderstands the relation between the desire and the good which Aristotle had already made perfectly clear. We desire things because they are good, and if we get the good thing, that makes us better off. But it does not make us better off because it fulfils our desires. In other words, desire-satisfaction is not a good-making property. That’s the difficulty. Whereas, something like accomplishing something with your life, that’s a plausible candidate for a good-making property. What the hedonist will say here is, “Well, these things that people say are valuable independently of pleasure—like accomplishment, friendship, knowledge and so on—they’re all things people enjoy”. So that already makes the objective list view slightly suspicious. But it’s not just that. Confronted with many common-sense principles that we’re taught—that you should keep your promises and so on—the utilitarian will say, “These principles are very productive of utility, so you should follow them”. It’s the very same with accomplishing something, having friends and knowledge and so on. These things will lead to pleasure. So the basic idea is that, just as someone like Sidgwick did, you can distance yourself from common sense both in your thinking about morality and in your thinking about well-being.

Now, if I understand your view correctly, you take pleasure to be a quale (2006a, p. 109). Which is to say, you take pleasure (or enjoyment) to be an element of phenomenal consciousness such that there is something that it is like to experience it. In short, you think that pleasure is a feeling.

Thus, if pleasure (or enjoyment) is a conscious feeling, and if hedonism is (in part) the view that only pleasures and pains count towards well-being, then this implies that non-conscious creatures cannot be welfare subjects. Would you agree?

Yes, I would. So, I’m putting what you might call a phenomenal experience requirement on well-being, or something like that—a consciousness requirement, maybe. But I know that some people like David Chalmers recently have taken seriously the idea that zombies could have well-being. And there does seem to be a link there with the desire and objective list theories. Because, on the face of it, you might think that zombies can have desires fulfilled and so on; they could accomplish something; and, on some conceptions of knowledge, they could know things, but it’s just that there is nothing it’s like to be a zombie. I guess someone like Chalmers would say I’m only holding this view because I’m a hedonist—but it seems to me totally implausible that zombies could have a well-being, that there is anything good or bad for a zombie.
This view has gathered momentum recently, and one of the arguments you referred to has come to be known as the argument from positive goods (Bradford, 2023). In a nutshell, (1) the experience machine shows us that there are non-experientialist goods, (2) these goods can be instantiated in the lives of non-sentient creatures such as zombies, so therefore, (3) non-sentient creatures such as zombies, robots, plants, etc., can be welfare subjects too.

Be that as it may, consider the following case. Imagine we stumble across a community of aliens, and they seem to care about others, and are cognitively sophisticated, and display a range of behaviours that are similar to those that we display, but it just so happens that they don't have phenomenal consciousness. On your view, they do not count as welfare subjects?

That's right. There's an epistemological problem about how we tell that these claims are true. But if the question is, if we were to do something apparently bad to those individuals, and it really were true that they did not have phenomenal consciousness, then I am committed to the idea that there is nothing wrong with our treating them in any way that we wanted to.

I do think that this discussion is really interesting. And, in a way, I think it's good for the hedonist, because it's produced a bullet for the objective list theorist to bite, one which hadn't really been noticed at the time that people like Griffin were writing.

What do you mean, exactly, by having a bullet to bite?

Well, you have to accept that a zombie who accomplishes something has made their life better for themselves.

That's very interesting. Because I feel as if it's the hedonist who is biting the bullet in these cases, especially in the example of the sophisticated aliens. I think you are quite right to point to the epistemological issue. Nevertheless, the hedonist is committed to the view that it would be perfectly fine to exterminate a race of sophisticated but non-sentient aliens. To my mind, that's a pretty big bullet.

Fair enough. It might be worth making a general point here about examples in ethics and other areas of philosophy. A lot of people seem to think that the way to do philosophy is to hear what somebody says, and then to come up with some consequences or implications of that account which look weird, and leave things there. I don't think that's a good way to do philosophy. Sidgwick is right when he claims that you need to get your views on anything in philosophy, including ethics, very, very clear—as clear as you can make them—and then you have sit down and think honestly about whether any particular view is self-evident, that is, necessarily true, true in all possible worlds. I think examples can be helpful here. For instance, say you do decide that pleasure is the only good, and then somebody comes up with an example like the zombie one. In that
case you should go back to your view, and you'll be able to see that it's
counterintuitive, but you need to ask whether it's so counterintuitive that
it undermines your initial confidence that this is a self-evident truth. And
if I do that, I find that I can retain my original confidence in hedonism.
Though of course I'm not saying that, in the case of somebody else, it
might not be different. And, Sidgwick was very good at explaining why
that should be worrying—i.e. that you and the other person both disa-
gree—unless you've got some account of why one of you is unjustified in
taking that different view.

Most people in this debate agree that sentience is sufficient for an indi-
vidual to count as a welfare subject, and I think this is consistent with
all that you've just said. This consideration seems to survive Sidgwick's
reflective process. The critical question, though, is whether or not it is
necessary for being a welfare subject—is sentience and the capacity to
have experiences of pleasure and pain the only thing that matters?
That's the crucial issue.
I think that's right. And of course, objective list theorists could refuse to
bite the bullet and say, “Oh, well, I have some kind of experience require-
ment”. But you cannot just bolt that on without thinking about it in iso-
lation, and asking yourself whether you are saying this just so you can
keep the show on the road.

Yes, and then, if they posit an experience requirement as necessary,
that raises the further question, why is that necessary? And hence we're
back at square one.
Yes, I think that's right.

Moving on, then, one objection to the view that pleasure is a quale is
the so-called 'heterogeneity objection', the idea being that various feel-
ings which we label 'pleasures' are so diverse that it is difficult to be-
lieve that they have anything in common whatsoever. For instance, the
pleasant feeling of a warm shower seems to have little—if anything—
in common with the pleasure I take in the fact that next week I'll be on
holiday.

Your response (2006a, Ch. 4; 2012) to this objection is to point
out that pleasurable feelings do have something in common, to wit, the
fact that they are all pleasurable!

But a sceptic may simply reply that we only label these feelings
as pleasures—that is, we use the word ‘pleasure’ to refer to them all,
but that this does not entail that they all share a certain phenomenol-
ogy. Rather—the sceptic might continue—we call these feelings pleas-
ures only because we take a certain pro-attitude towards them. Why
should we believe the former view and not the latter?

3 See, in particular, Feldman (2004).
Because there is a difference between a feeling and an attitude. Pleasure is a feeling. And we can now measure it. There are certain circuits in the brain which are associated with the feeling of pleasure. And you could have one of these pro-attitudes without feeling pleasure. So, for example, if we think about Feldman's idea of attitudinal pleasures, you might say, “I take pleasure in and am pleased at the fact that the Nazis lost the war”. But there might be nothing going on in the pleasure circuits in the brain. You just think it's a good thing. You don't actually have any feelings about this. You could have feelings. You could contemplate that, and have certain subsequent feelings as a result. But it's still the case that your taking pleasure in the having of the attitude. It's not the attitude itself that is the pleasure. I think it's fine if you want stipulate that that's what you mean by pleasure—i.e. that it's an attitude—but I don't think it's in line with the hedonist tradition, or indeed ordinary usage. And to me it doesn't look very plausible as the foundation for an evaluative position. I can't see what's so great about tokening an attitude if you're not getting anything pleasurable—i.e. any enjoyment—out of it.

A Recent paper, Lin (2020) argues that pleasure is an attitude, but goes on to ask, if we agree on this much, then how do we identify this attitude—how do we bracket it off from other attitudes that we have? Well, Lin thinks that the way that we do this is by identifying those pleasurable attitudes with the ones that feel good. Thus, on this account, attitudinal and phenomenal hedonism don't conflict. And hence we have a hybrid account of pleasure.

I partly agree. An attitude you take pleasure in is a hybrid. But it doesn't mean we've found some kind of new philosophical pigeonhole. I think that the distinction between attitudes and pleasures is still there. It's just that they can be found together.

More recently, there has been a growing body of work discussing the badness of pain.4 Two cases are often discussed in the literature. The first is pain asymbolia. These are cases where individuals report feeling pain, but do not seem to be bothered by this—i.e. they have no con-attitude towards it. The second type of cases are what has come to be known as 'Hurts-so-good' cases (HSGs). These are instances such as climbing a mountain, having a deep-tissue massage, or jumping in a freezing lake, in which people tend to enjoy pain for its own sake, and so, paradoxically, pain seems beneficial in HSG cases.

Now, I am making the following assumptions. If you take pleasure to be a feeling, then, presumably you think that there is some essential feature of this feeling that makes it intrinsically good. And, assuming that pain is the opposite of pleasure, then I suppose you take some essential feature of the feeling of pain as intrinsically bad. In the

---

4 See, for example, “Well-being and Ill-being”, Midwest Studies in Philosophy, Vol. 46. 2022.
two cases given above, however, there appears to be nothing intrinsically bad about having these painful experiences. They appear to show that pain is not itself intrinsically bad, and hence they pose counterexamples to your view. How do you respond to these counterexamples?

I think these cases are fascinating. And I’ve had some pretty interesting discussions with people about them, including the neuroscientist Morten Kringelbach. Now Morten doubts them because he thinks that people can just be misremembering, misinterpreting, misunderstanding their own experience. And as he says, we’re actually quite good at that as a species. But I think another possibility is that, there is, as it were, a sensation of pain—there is something that it is like to be in pain—and maybe that’s the same before and after the operation or whatever, but that’s not what hedonists are talking about. Hedonists are talking about the experience of having that sensation and disliking it. But you have to be careful. It’s not that the dislike, as it were, is making the pain bad. Because then that would make other things bad, even if they were completely independent of pleasure and pain. The idea is, for beings like us, having sensations of pain and sensations of pleasure and disliking them in those cases is a pretty basic category of experience. And that I think would deal with the hurt-so-good cases.

Right, so to get clear on this, you’re saying that there is the sensation of pain, but I like this sensation of pain, and so I’m enjoying it. And, on your view, is this good or bad for a person?

It’s good for a person—it’s good if they’re enjoying that sensation of pain. This is what hedonist meant by enjoyment. And, on this view, all enjoyable experiences are intrinsically good. That’s why people jump in the lake.

But this seems fairly attitudinal. I have the feeling of pain, but my attitude of enjoyment is what makes it good?

It is attitudinal because it involves attitudes. But the attitude isn’t making things good or bad. And it is complicated. For example, I sometimes go swimming in the river near my house when it’s quite cold, and when I jump in it’s a mixed experience, there can be a certain amount of pain involved, so there are bits of it I don’t enjoy—I dislike them. But then, as I get more used to it, I start to enjoy it. But I think there are purer cases, where the person, right from the start, is feeling the sensation of pain and they’re just loving it. And that’s taking pleasure in the sensation of pain.

But there has to be a feeling of pleasure.

There is the feeling of enjoyment, and that’s just the experience of liking the pain.

I think I see it now. So, there is a feeling—something that it is like—to like/enjoy the sensation of a pain. So there is an attitude involved, but that attitude is accompanied by a phenomenology, and it’s the latter phenomenology and not the attitude that counts?
Yes.

*Hedonism is often criticised on the basis of the philosophy of swine objection. Very roughly, the idea is that, if pleasure and pain are the only things that matter, then the life of a contented pig can be better than that of a discontented Socrates, as Mill (1863, Ch. 2) puts it. This is a proposition which many people find difficult to believe.*

The implication of the philosophy of swine objection is that there must be some point at which the prolonged simple pleasures outweigh the rich and diverse pleasures of human experience. So, for example, if you give the pig enough of its piggy-pleasures, then there will always come a point at which the pig's life is better than that of the human's. You have responded (2006a, Ch. 4) to this objection in the following way.

You claim that deciding which are the correct criteria for valuing enjoyment (pleasure) should be influenced by the way subjects actually value experiences. So imagine an individual judges reading a novel more enjoyable than drinking a glass of lemonade. If I understand you correctly, you claim that, so long as there are certain properties of the experience of reading the novel which the subject deems good, then we do not have to assume that there must exist an amount of prolonged lemonade-drinking enjoyment which necessarily outweighs the enjoyment of reading the novel. In other words, this valuation of the hedonistic value of novel-reading could be correct.

This is a complicated view. And so, before we move on, I would like to invite you to say a bit more by way of clarification, and why you think that this is the case.

Yes, I think what I was doing there, taking my view from Griffin and others—maybe Rashdall—was suggesting that it’s possible, not totally implausible, that there are discontinuities here. So, yes, it’s true that the more years of piggy-pleasure you get, the better off you are. But it does not matter how many years of piggy-pleasure you get, it’ll never compensate for (let’s say) a year of higher, human pleasures. And it could be the same with pains. So you could have what I call *bad-bad* discontinuities, as well, in which, for example, if you take a week of the most awful agony you can imagine, say torture, and compare it with a lesser bad, say, a hangnail, then, it doesn’t matter how long you have the hangnail, it’ll never be as bad as the week of torture. Now, that seems to me an available position, and one that’s not completely implausible.

*But apart from that general intuition, is there further argumentation to back this position up?*

No, but that’s OK, because I think we’re getting down to bedrock at this point. I’m a foundationalist on these matters.

**IV. PRACTICAL REASON**
Turning now to normative considerations. W. D. Ross entitled his seminal work, *The Right and the Good*. The title of your book, *Reasons and the Good*, subtly implies that this distinction misses the mark.

This is because one of the key theses of the book is that we can answer—in full—the fundamental normative question, “What ought I to do?” without reference to moral concepts. As I understand it, all that is required on your view is the concept of (i) an individual’s welfare, (ii) normative reasons, and (iii) a mechanism for comparing these reasons, such as ‘weighing’.

*Is this correct, and would you care to elaborate at all?*

I think that’s absolutely spot on. That is the view I ended up with, but I didn’t start with it. The road I went down could have led to returning to the ‘right’ as well.

I think it’s amazing in philosophy how rare it is for anyone in any area of philosophy to start by saying, “This is the question that I’m asking. And I’m going to stick to it, and I’m going to give you the answer”. So in general this simple, basic principle of good philosophical methodology is widely ignored. So which question are we asking in ethics? Well, you might say, “Oh, it’s obvious, we’re asking about what the right thing to do is”. Well, people are doing that. But if we get the correct answer to that question, there’s still a prior question, viz: “do we have any reason to do what’s right?” I think this suggests that the ultimate question in ethics is what reasons we have.

What is more, one should be parsimonious. The philosophical conceptual scheme we have is, on the whole, just given to us. And we often have lots of concepts which roughly amount to the same thing. They’re often not quite the same, and one can go on for ages about how they differ from one another. But it gets very complicated, and there’s a lot of quicksand there that it would be good to avoid. So, I think that, once you’ve decided what your question is, then you need to start answering it. And every time you offer an answer, using certain concepts, you should ask, “Do I really need that concept? Because, if I don’t, I’m just going to put it to one side. Maybe it’s unnecessary and maybe it will cause confusion”. So there’s an argument for parsimony here.

Again, following through on Sidgwick’s view, I don’t think you have to go that far, ending up by saying the reasons we have to act are to do with the promotion of net well-being, that is, net pleasure, and nothing else. As Sidgwick himself points out, we’re separate beings. And the egoistic view, that you have strongest reason, or reason only, to look after your own good, has been viewed as very plausible over the centuries. Indeed William Frankena once claimed it’s probably been the most popular view over time. (This is another strange thing that happened in the 19th century: people largely just stopped talking about egoism. And there was very little discussion of egoism in the 20th century.) Anyway, the point is that there are several views on the table: hedonistic (perhaps weighted) utilitarianism, hedonistic egoism, or some dualistic view, according to which what we have strongest reason to do in any case depends on which
principle speaks more strongly in favour of acting in accordance with it on that occasion.

As an aside, as I view 20th-century philosophy, it made some massive advances over previous centuries, in clarifying problems and possible solutions to them, but, in general, certainly in ethics, things really went downhill after Sidgwick. People on the whole didn’t recognize the power of his arguments for utilitarianism and for egoism, as well as the problems he found in deontology. (I think that book III of Sidgwick’s Methods, along with Richard Price’s Review, are the best statements of deontological ethics in the history of Western philosophy, at least since Aristotle. Further, not only did Sidgwick state the views pretty clearly, but he also had some quite powerful objections which people didn’t bother to deal with. W. D. Ross is an exception. I think that Ross was a bright point in 20th-century philosophy because he basically saw the world in a very similar way to Sidgwick. He was an intuitionist. And he was very careful in his philosophy. But he also saw the plausibility in the deontological view which Sidgwick hadn’t completely dispelled. And that was the thing to think about. But unfortunately people went off in the direction of metaphysical meta-ethics. And then, in the 70s, for understandable reasons, practical ethics. That’s worth doing, but if you’re going to do practical ethics properly, then you need to have a theory, and we haven’t made much progress on that since Sidgwick and Ross. It’s nearly a hundred years since the Right and the Good, so I hope things improve.

Anyway, I’ve gone slightly off track. On the demoralization position, it’s based on parsimony, but also, I think, there are two further broad kinds of argument. One would be that, when we think about morality properly, we can see that it’s evolved. It’s a cultural and social phenomenon with a history. Why have these systems evolved? Because they promote survival, broadly understood. Now, if you think about the law, things have changed. So, in the past, many people were natural lawyers. They believed that there were positive legal systems, but also that, on top of this, there was a natural law built into the universe, decreed by God or whatever, so that we can consider positive legal systems in light of what we think about natural law. These days, most legal theorists have given up on that way of thinking. But, in ethics, people continue to think of morality in an analogous way, as if there’s some kind of ideal moral law built into the universe, and we can find out what it is, and we can then test our own moral systems against that. But I think positivists in law are correct, and that we should be positivists about morality. And we should probably be positivists about religion, at least to start with. This is not to say that we shouldn’t be doing philosophy of religion, but that we should hold these positions at the start. And in ethics we should ask whether we actually want to go along with the idea that there’s this—on the face of it, rather peculiar, and hard to explain—moral law, in the universe, though we don’t believe in God anymore, demanding that we do certain things, requiring that we do certain things, requiring that we punish people
because they’re blameworthy, and so on. No, let’s not do that unless we have to.

The other worry is that, more than in the case of law, morality involves the emotions. Many people—particularly outside of philosophy—when they are having moral arguments become very worked up. It does happen within philosophy as well—there is too much table thumping going on, too many high-minded appeals to ideas of nobility, dignity, and so on. This is not a good way to do philosophy. You should be standing back from your emotions, and using your capacity to reason; and, as I said before, working the way that Descartes and Sidgwick and other people have recommended: state your view as clearly as possible and as unemotionally as you can, and then decide on the basis of calm reflection whether it’s true. Now that’s not to say that you might not need certain emotions, or the capacity for them, to grasp necessarily true practical principles. It’s not a coincidence that many utilitarians were highly benevolent, and they probably wouldn’t have got to that view if they hadn’t been.

According to your view, deontic concepts such as right, obligation, duty, etc., fall out of the picture. And the question, “What ought I to do?” can be fruitfully rephrased as, “What do I have most reason to do?”. Ross, therefore, had no need to appeal to the deontic concept of ‘right’ when ‘most reason’ would have done just fine.

I must admit, I find this view very attractive. Let me ask, therefore, what do you consider to be its main downsides?

Well, one downside is that many people seem to disagree with it. Like Sidgwick, I’m attracted to Pyrrhonism, and this means that, strictly speaking, I have to suspend judgement on several views which seem to me to be highly plausible. I have had conversations with people about it, and it just is true that some obviously intelligent and reasonable people will say, “I’ve thought hard about the ultimate question in ethics, and I think it’s ‘What’s the right thing to do?’”. I might reply “Well, but do you have reason to do the right thing?” And they’ll say, “Yes, of course you do, because we have the correct conception of rightness as itself grounding reasons.”

So there are certain conceptual debates that I think still need sorting out. And I guess a practical downside of the view is that you can’t engage straightforwardly in the kind of praise and blame that many people believe belong to morality. I think that in *Reasons and the Good* I used the example of Hitler. We want to be able to condemn Hitler in the strongest terms. We want to be able to say that he was a monster, evil, contemptible, and so on. We want to use all these moral terms to condemn him. But according to the reasons-first approach, or whatever you want to call it, all you can say is that Hitler had very strong reason not to do what he did. I think the only solution here is to recognize the two levels again. So demoralizing on the whole is a philosophical strategy, but I also think that in certain cases it can be quite helpful. Especially if tempers are getting heated and there is a lot of disagreement going on. It can be valuable for
people to step back and recognize that what really matters in the question at stake is how certain lives go.

*Your view in Reasons and the Good is teleological—since all reasons for action are those that promote welfare. But you state that the view is not ‘act utilitarian’. Presumably this is for two reasons. The first being that the view takes into account the distribution of welfare.*

*But, second, the view places more weight upon self-interested reason than does traditional act consequentialism. You call this the ‘dual-source’ view. Very roughly, the idea is that there are reasons to promote both one’s own welfare, as well as the general welfare of everybody (see also, Crisp 1996a).*

As you note, the dual-source view relies on the idea of the separateness of persons. But the separateness of persons can be, and has been, challenged. *Why are you not swayed by these arguments?*

The most significant aspect of Parfit’s (1986, part III) work is that he’s brought out more clearly than anybody that I know of in the history of philosophy just how important the issue of personal identity is when it comes to ethics. In ethics it’s important to find out, as it were, which beings we should care about (some people speak in terms of moral status, but that’s a bit misleading), i.e., which are the beings that matter. Does that chair matter? Does that bird matter? Well, the bird matters but the chair doesn’t. But when it comes to human beings, there are some real problems about the correct view of personal identity. So if you’re going to hold the dual-source view, you will have to get clear on that. And I haven’t, really, yet anyway. But say that one followed through on Parfit’s sceptical arguments about the standard views of personal identity all the way to Buddhism of a kind, where you’ve got just momentary subjects of experience. Even then, on the face of it, they would look to be separate from one another as well. Now, you might just bundle them all together and say there’s just consciousness, and a single subject. Then the dual-source view would in effect collapse because there would be no duality. But until we get to that point, I think there’s going to be separateness.

So, why do I accept the principle of self-interest? Well, it's partly through noticing that everybody that I know thinks their own self-interest matters to them in some special way. But, of course, that doesn’t show that it does. It's more a matter of reflective intuition, where you do have to focus on particular cases. Imagine one case where you’ve got to decide whether to undergo some pretty bad pain and suffering. And if you don’t do that, some very minor hangnail-type suffering will occur to many other people. But these people are already doing massively better than you. I’m not saying you shouldn’t undergo the severe pain, but I just cannot imagine how it could be at all plausible to say that the fact that it’s *me* who’s

---

5 Crisp holds a ‘sufficiency’ view. Namely, that we have special reasons to promote an individual’s welfare up until a certain threshold, after which, these reasons give out, and we have no (non-derivative) reason to give any individual special treatment (see 2003; 2006a, Ch. 6).
going to be experiencing this severe pain, and the fact that they’re going to be experiencing the hangnail, are just practically irrelevant. I just cannot imagine thinking that all that matters is the impersonal pleasure and pain. That’s where I get off the bus really, and I do wonder if utilitarians might be charged with at least a degree of bad faith at this point.

There does seem to be an intricate connection between the entity/subject that’s making the decision and the experience of the same entity/subject that’s making that decision. It seems that because the entity/subject that makes the decision is the same entity/subject that will have that experience, then this provides that decision-making entity/subject with reasons that other entities/subjects who will not have that experience do not have.

Exactly.

So, such a case is difficult to imagine, but we can try to imagine an angel version of yourself that makes all of your decisions from heaven, but your earthly body is the one that has to experience the effects of the decisions taken by your angel-self up in heaven. And thus, in this case, the subject of experience and the decision-making subject are separated, and thus the pain of the one seems less relevant to the other. (Notwithstanding the significant difficulties concerning the theory of personal identity here.)

That’s a nice example. My main problem with it is understanding the identity of the angel with the body. It sounds as if this is a case of one person’s making a decision about another. But I’m not ruling out that claim of identity as wrong or inconceivable. Many years ago, for example, people would have found it inconceivable that parallel lives could meet. But we now know that, given the way space actually is, they do. So I can imagine arriving in heaven and God’s saying, “That angel is identical with you”, or perhaps it could turn out that utilitarianism is in effect true, because there is just one single consciousness, and ‘I’ am just a part of God’s consciousness.

One line of thought in Parfit that I don’t think he followed through properly, and I don’t think reductionists have followed through properly either, is this. If you take Parfit’s case of My Division (1986, pp. 254–255) and like Parfit you think that we split, in the Y case, if you think that’s twice as good as ordinary survival, what you have to accept is that your well-being is now being accrued in two different places at the same time. So, if you could come up with some view in which all sentient beings were identical to one another—they were the same owners of well-being—you could plausibly become a utilitarian. But it would require a very peculiar metaphysics, I think.

Of course. We would have to fundamentally rethink what it means to respond to reasons and make decisions in such a case.
On your view, then, the reason that my welfare is my own gives me a reason for promoting it over another's provided that the costs to me are significant enough and the benefit to the other is not very significant. So, in a type of case that you discuss (2006a, Ch. 5, part 2), if I could make it so that either I, or someone else, receives an electric shock of the same magnitude, then, all else equal, I have most reason to prefer that you, rather than I, receive the electric shock.

Personally, I find this difficult to believe. I struggle to see how—on a view in which all normative reasons are derived from welfare-promotion—the mere fact that this welfare is mine (and not another's) gives me a normative reason to promote it—at least in some marginal cases.

I agree that this is intuitive. But to my mind, we can explain this intuition away. The reason why it may seem as if I have most reason to promote my own well-being in this case is because (generally speaking) I will also have motivational reasons to promote my own well-being, too. So while I have equal normative reason to shock either one of us, my motivational reasons tip the balance—they act as a tiebreaker as it were.

Hopefully I have made this point clear. To try to put it differently, the reason it appears intuitive that we have most reason to shock the other person is because we wind up smuggling in motivational reasons into our normative analysis. If what I say is correct, however, your view ends up implying that, just because we are more motivated to shock the other person, then this is what the balance of reasons tells us to do.

Surely you reject this analysis. So why is it incorrect?

That is a worrying point. Another case might be one in which you and I travel to some planet where we find beings like us who are highly rational and intelligent. But they're completely impartial. I would then have to say to them, “Actually you're making a mistake, because you haven't taken the separateness of persons into account”. They would then, like you, say that, we know about evolution on your planet, and you've just evolved to be partial, but this isn't really reasonable.

So I agree. There is a powerful debunking argument there. But there are no knock-down debunking arguments of that kind. I think they're good challenges. But what you have to do is go back to your original view and ask yourself whether you can still accept it in light of the debunking argument. And I can still accept it, because of cases like the one I mentioned before about pain, and so on.

Essentially, I'm following a tradition which goes back a long way, maybe to Plato, if not before, according to which ethics is like mathematics. So in ethics what we're trying to do is to come up with necessary truths, propositions true in all possible worlds. So if I say that 7 + 5 = 12, and someone says, “You only believe that because we evolved to accept useful mathematical principles”, I would reply, “Yes, well, that happened, but it just is the case that 7 + 5 = 12”. I would say the same about pain's being bad. We've obviously evolved to believe pain to be bad. But it's bad
in all possible worlds. And one can see that—at least, one *should* be able to see that if one reflects properly on what it is to suffer.

**In your book, you concede that self-sacrifice can be irrational. To soften the blow, you offer two arguments to support this conclusion. You first point out that, in extreme cases, this seems right: for example, it might be irrational for me to sacrifice my life merely so you can have another cup of coffee. Secondly, you highlight that the term ‘irrational’ is an emotionally laden one, giving rise to connotations of blame and so forth. Hence it is better to use phrases such as “most reason” instead (p. 133–4).**

Still, this is another implication I find difficult to swallow. To see this, let us adjust the preceding example slightly.

Suppose that I am a very altruistic person. I have a very strong desire that I receive the shock instead of you. Now suppose that, were you to receive the shock, the net loss to your welfare would be ever so slightly worse than if I received the shock (adjusted for any counterbalancing benefit I may receive from having this desire satisfied). Suppose that I know all of the facts—yesterday I received a shock of the exact same magnitude, so I know exactly what it feels like. I am perfectly lucid and thinking clearly. Nonetheless, I still really want that I (and not you) receive the shock. When I make it the case that I (and not you) receive the shock, I am not acting according to what I have most reason to do—I am acting contrary to my reasons.

This seems like an alienating notion of normativity. How, when I am perfectly lucid, know all of the facts, and act in accordance with what I most want to do, can I still come out as acting contrary to reason in such a case?

**Perhaps this comes down to a deep disagreement between intuitions, here. But why do think I might be wrong?** That is another problematic implication of my view. I would try to deal with it in the same way that I did with the previous one. Again, what we should say about these cases depends on the correct account of personal identity. But let’s just assume that the natural view that we have is correct—that each individual is distinct—then, I can see the point of the case and it worries me. But my original intuition—which Sidgwick has as well—is that the distinction between persons is not just metaphysically important, but it’s normatively important. And so I’d stick with it nonetheless. And also, I can again appeal to the two-level view from earlier, in order to explain why we have these attitudes. Because I obviously have the attitude prompted by your case as well. I think that this altruistic person is admirable. But that is just because we’ve been brought up with the morality that we have.

**Sure, but the point that’s worrying is that there is a top-down, alienating requirement in this case. Since in the example, I am perfectly lucid,**
I know what the pain will feel like, and so on. So it seems as if I am acting rationally, here. Yet on your view I'm not acting rationally. That is my view. But, in a sense, I am only reporting it to you. I am not reporting it to you and banging the table, because I have no justification in doing that. That's the way in which philosophy should proceed. People should simply be reporting their views for others to criticize them. And so there are lessons in the Sidgwickian tradition for how we should go about doing philosophy—undogmatically, open-mindedly, calmly, and carefully.

Now I'm thinking out loud here, but it could be possible that what's doing the work in this example is that it is vague. What I mean by this is as follows.

In the extreme case, it does seem that I'm being irrational when I prefer the huge electric shock for myself as opposed to the tiny shock to you. Likewise for the other extreme: I have more reason to prefer a tiny electric shock for myself rather than to impose a huge electric shock upon you. But when it comes to marginal differences in welfare, things are more tricky. It would appear as if your view needs to assume implicit thresholds as to when it is irrational for me to prefer the shock to myself as opposed to you. In the extreme cases, we can easily intuit when the threshold has been passed. But in the marginal cases, intuitions differ. And that's where the difference in agreement creeps in.

Yes, that puts the point very well. And these are exactly the kind of cases we should be focusing on, rather than the dramatic ones. It is also particularly worrying because, like Sidgwick, you've got a very partial principle, self-interest, and a very impartial principle, the utilitarian one. And many counterexamples to egoism and to utilitarianism rely on the idea that we have special reason to favour the interest of those who are close to us: family friends, and so on. I tested my intuitions about that, and came out thinking that Sidgwick was right. But equally I'm pretty sure that the majority of people would come out thinking that he was wrong.

V. LOOKING BACK (AND FORWARD)

Out of the philosophical problems that you've attempted to deal with, which attempt are you most dissatisfied with?

I've never considered this question before. And it has got me thinking about the attempts with which I'm most satisfied. For example, I think my argument against the coherence of virtue ethics is pretty undeniable. I haven't come across a good response to it. And I can't see how there could be one.

Hedonism—I'm pretty satisfied with that. Though here I can conceive that there might be some arguments which would persuade me to give it up.
What I'm most dissatisfied with, I think, would be the issue of personal identity, I'm still trying to get to grips with that. Ingmar Person has some superb stuff on this, where he suggests that our views about personal identity are so confused that we can't make any progress. I hope that isn't true. At least, I hope people could make progress on this in the future even if we can't make progress on it now.

I am still slightly worried about the issue that we were talking about before, namely, impartiality towards others. Because we’re so limited in our self-knowledge, it could be, for example, that something terrible has happened to me that's made me deny the importance of partiality to family and friends, that's made me put too much weight on the idea of impartiality.

**Which philosophical problems do you still find most fascinating?**

I suppose at the moment, those two, personal identity and (im)partiality.

I have a friend, Paul Robinson, a very good philosopher, who took the B.Phil. at the same time as me. He continued with philosophy and started a career in philosophy. But then he switched to another career. I was talking to him about why he did that, and he said, “I went into philosophy to answer certain questions for myself. I’ve now done that. So I’m going to do something else”. I haven’t got to that stage yet, and there are lots of other things I’d like to do that I wouldn’t mind doing. I would be quite pleased if I could get to the point Paul reached.

I will always be interested in problems in metaphysics. But I wouldn’t want to think about those only within philosophy. I’d want to think about them alongside scientific evidence as well. Mainly personal identity, but also more fundamental questions about the nature of reality. I suppose I’m still interested in the philosophy of religion, and I taught it for a long time. I do still find the cosmological argument quite plausible. (Derek Parfit used to get very cross whenever I told him that!) But that would tie up with general thinking about cosmology.

In epistemology, I’m happy as a foundationalist. I’m not saying that the arguments I have for foundationalism are knockdown, but they’re convincing enough for me.

**What are your research plans for the future?**

Well, I hope to think more about the above. In the shorter term, I’d like to think more about a related set of issues to do with how we should think about ethics.

Like a lot of people, I started with a rather egocentric question: “How should I act?” I don’t think I’d taken stock of the fact that even just seeking to answer that question is itself an action. So that raises the further question: “How should I seek to answer practical questions?” My view is that the consequences of our all asking that initial egocentric question are catastrophic, and will be even more so in the future, sadly. Take something like climate change: this is the result of self-interest combined with people asking the “I” question, and then continuing to fly and so on when
they realise that what they do is not going to make any difference. It's true that if we started thinking as a group and decided that we ought to stop flying, someone might say: “OK, but then you're going to have to ask the question: “Should I fly?” But we shouldn't be thinking about the “I” question at all. We should just be thinking about us. So impartiality about the fundamental ethical question is, I think, really important practically speaking. But I also think it's theoretically interesting in ethics. And relatively, I think there are arguments about meaning in life, which relate to this. Most people seem to think that well-being and the meaning of life are independent of one another. So they might say that my life had a lot of meaning but it had little well-being. You might imagine someone’s saying this about Wittgenstein's life: highly meaningful—as my friend Dave Edmonds once put it, it contributed to the CV of humanity; but perhaps it wasn't so great for him. So meaning and well-being might be different. And there may then be arguments for thinking about the way one lives one’s life which are independent of what Frank Jackson (1987) once called the ‘difference principle’—the principle that the justification for acting in some way must depend to some extent on your making a real difference. But think about for example people who played a part defeating the Nazis. If you talk to people who did that, and played a non-trivial part in it, it clearly made a big difference to their life. But it would be absurd to think that in the possible world in which, say, they broke their leg, and couldn't go to Dunkirk, everything would have changed. So what we do, if it makes a real difference, seems to make our lives meaningful. Recognizing that could perhaps make a practical difference and help us solve some of the various and apparently intractable collective action problems.

I'm also working on a translation and commentary for OUP on some dialogues of Plato: Euthyphro, Apology and Crito. I've completed the Euthyphro, and am just getting going on the Apology—it's not going very fast, but I hope to get there in the end.

And then, in the longer term, I'd quite like to write a book on Mandeville's ethics. I think he has some views that are really interesting and suggestive. He's become almost a forgotten figure in ethics, and I think that's because people focus too much on Hume. There is much in Mandeville that we'd do well to remember.

**How would you reflect on your philosophical career to date? What are you most grateful for? And what would like to have done differently?**

When I look back, I think I can see some kind of narrative arc in my career. I seem to have been on the whole rather negative, sceptical, unfashionable, and contrarian. I've argued among other things that everyday moral terminology should be avoided in philosophy; partiality to others is ultimately irrational; practical reason is disunified; the idea of supererogation is incoherent; hard incompatibilism is correct; there is no such thing as 'virtue ethics'; equality is if anything a disvalue; hedonism is true; we should suspend judgement in ethics; national borders, and hence nations, of the kind that exist currently are unjustifiable; differences between our
morality and that in Homer are of degree rather than in kind; Bentham and Mill are intuitionists, prefiguring Sidgwick and Ross, and there has been little progress in ethics since Ross; most advertising should be banned; it is not unreasonable to think that the end of sentient life on earth would be good; and so on ...

In general, I feel as I have been, as it were, on the outside of morality, and moral philosophy, finding kernels of truth in both but rejecting a good deal (this is what I think of broadly as ‘demoralizing’). I’ve turned out to be a bit of an intellectual loner, and this is not, I think, what I expected. I do sense among some younger philosophers a greater willingness to think more radically and abstractly about ethics, and this gives me hope that we might make some progress beyond our current impasse.

As far as gratitude is concerned, I think all of us who have had lives which are overall better than nothing might reasonably be quite grateful (or glad) if we think about the incredibly small probability that we were going to be born. I’ve been extremely lucky to live at the time I’m living in, in the places I’ve lived in, with the parents, family, friends, colleagues, and students I have had, the school I went to, and much else. As I’ve mentioned, when I started in philosophy, I wasn’t too worried about whether what I said was true or not, but just wanted to come up with an argument that got me a high mark. Perhaps I was really just trying things out. Later on I started thinking that I would like to come up with some answers to philosophical questions, and then I went through a period thinking this wasn’t going to happen. Just because it didn’t. Then it did start to happen. I did start to come up with views which I thought were correct. But then I became a Pyrrhonist. But that’s not the same as the earlier failure. Becoming a Pyrrhonist I see as a kind of success. And it’s probably made me more relaxed psychologically. I don’t become impatient or irritated with people who hold views different from mine in the sense that some other philosophers seem to.

What would I like to have done differently? Not much. I think I might have become a classicist, or perhaps a historian, but am quite happy as a philosopher.

VI. ADVICE FOR GRADUATE STUDENTS IN MORAL PHILOSOPHY

*If you had to recommend, say, three or four texts that you believe any graduate student in moral philosophy should read, what would they be?*


*What other advice most helped you as a graduate student?*

I mentioned John Ackrill before, and in my first meeting with him, he gave me two bits of advice, which presumably he gave to everybody at that time. The first was learn to type, which, back in 1983, most of us couldn’t
do. That was very far-sighted. I took his suggestion and it was fantastically helpful in all sorts of ways. So I would say to students: be very aware of the technology that’s available to you, and use it.

The other thing Ackrill said to me was you’ve got to be ready to talk to people. You’ve now got this opportunity to engage in discussion with others who are interested in the things that you’re interested in, and you’ll learn only through talking. Force yourself, if you have to, but it will become second nature. Obviously there’s an Aristotelian mean here, as there are plenty of graduate students who are too happy to talk. You’ve got to be prepared to listen as well.

One other thing. During the B.Phil., I would just get up in the morning and start reading. And apart from engaging in some social activities and so on, I’d then go to bed. And I did that for two solid years. Alan Ryan said to me that’s the way to do it, because you’ll never have a time like this in your life again. Read everything you can. Think about it, try to remember it, take notes. That was such good advice. I still have all of my notes from that period. Ryan told me both that I’d learn a lot from creating these notes, and be able to use them for at least the next 20 years. And he was absolutely right.

What advice would you offer to a young scholar hoping to embark upon a career in moral philosophy?
I would say that our culture is (slightly worryingly) achievement-oriented—I see this a lot in Oxford, obviously, where achievement matters a great deal. But I think it’s wrong just to assume that someone who has been successful in some profession has had a life that’s good for them—I think that’s a big mistake. It’s fine to go into that profession, but you should be really careful to go into something you’re attracted to for its own merits. So, if you go into philosophy because you think that your parents want you to, or you think it’s cool, or just because you have a certain natural talent, or whatever, don’t do that. Start with the question, “Am I going to enjoy it?”. I’d also advise this young scholar to talk to people who are doing it and ask them whether they enjoy it, and if so, why.

Which advice have you offered graduate students in the past but no longer do so?
I used to tell graduate students what I’d been told: don’t worry about publishing. But now things have changed, through greater ‘professionalization’ of the discipline, research assessment exercises, and so on. So now my advice is: “Publish if you like. But be careful about it. Don’t rush into it, and later regret what you’ve published and the time you’ve put into it”. Of course if one’s hoping for a postdoc., publications are very likely to be necessary for one to be considered.

One other thing. I would also encourage graduate students to be ‘good citizens’. This will give you an insight into what it’s like to be an academic. But also, it does strengthen your profile: if you’re going for a
postdoc position, say, maybe that will require some non-trivial admin work. If you’ve never done any, then that’s obviously going to count against you.

**As a graduate student, did you have a plan B in case you didn’t make it as an academic?**

I was thinking about law, and I did some mini-pupillages at the bar. I suspect that, given that I loved classics so much, I might would have tried to be a classics teacher. And I think I would have enjoyed that. Not so sure about the law...

**Rejection and failure often play a large role in the life of an early career scholar. What were your biggest failures and how did you overcome them?**

Well, I applied for a lot of jobs. You just have to keep going. And of course I’ve lost count of the number of times I’ve received rejections from journals. In the days when we submitted papers by post, Derek Parfit used to advise students, when they sent a paper out to one journal, to prepare an envelope for the next one. Try to see it as a learning experience: if you get a rejection with some critical comments, take a deep breath and read them, carefully.

**REFERENCES**


What Public Policy Can Be: An Interview with Matthew Adler


The Erasmus Journal for Philosophy and Economics (EJPE) interviewed Adler about his formative years (section I); his work on the theoretical foundations of public policy, zooming in on welfare-consequentialism and social welfare functions (section II), welfarism and interpersonal comparisons (section III), the ethical deliberator and the role of the philosopher (section IV); and, finally, his views and visions for interdisciplinary

---

1 An edited volume based on the work of this network has just appeared. See Adler and Norheim (2022).

EJPE’S NOTE: This interview was conducted by Måns Abrahamson and Akshath Jitendranath, editors at the EJPE.

Footnotes throughout the interview are editorial annotations directing readers to the relevant literature discussed in the text.
work in law, economics, and philosophy, as well as his advice for graduate students in the field (section V).

I. FORMATIVE YEARS

EJPE: Professor Adler, perhaps you can start by saying something about the trajectory of your career as a scholar of Law and Economics and the events, people, or writings that have had a particular influence on the development of your interests.

MATTHEW ADLER: I was actually a history major undergrad and then ended up at Yale Law School, where I graduated in 1991. Yale Law is a great place in terms of Law and Economics. For example, Guido Calabresi taught there for a very long time. I ended up taking more classes in constitutional theory than in Law and Economics—Yale Law is also very important in constitutional theory. So, although I had some exposure to Law and Economics at Yale Law School, my serious interest didn’t really blossom until after I started law teaching at University of Pennsylvania Law School. And I was lucky to have Eric Posner as a colleague there at the time, who had started a year before me. We started thinking about cost-benefit analysis, had conversations, and then wrote an article that came out in 1999 and that was the basis for the 2006 book, *New Foundations of Cost-Benefit Analysis*, where we challenged the traditional Kaldor-Hicks view of the justification for cost-benefit analysis.² So that was a formative series of events.

Then I shifted from cost-benefit analysis to social welfare functions because it seemed to me that we could do even better than cost-benefit analysis in terms of tracking social welfare. And at that point I was lucky to have another colleague, Chris Sanchirico, who taught me a lot. He’s a PhD economist and had been on the economics faculty at Columbia University. He had a big influence in terms of my understanding of Law and Economics. And since then I have been lucky to collaborate with a lot of people. For example, I encountered Marc Fleurbaey, I think, maybe in 2008 or 2009, when he came to a conference at UPenn and I have since worked with him on various projects.³

---
³ Adler and Fleurbaey co-edited *The Oxford Handbook of Well-Being and Public Policy* (2016); other joint work includes Adler and Fleurbaey (2018a, 2018b), Adler et al. (2020), and Adler et al. (2023).
The other thing I should say is that I started teaching at the University of Pennsylvania Law School in ‘95, which at that time had (and still does have) a very strong legal philosophy group. Not a philosophy of economics group, but a legal philosophy group. And a lot of my interest in philosophy—and knowledge, really—started with conversations with those colleagues.

As a follow-up question for clarification: Was the challenge to the Kaldor-Hicks criterion that it is insufficient as a moral criterion to justify cost-benefit analysis?4

Exactly. In the United States there’s a weird disjunction between the view of cost-benefit analysis in applied and in theoretical welfare economics. Kaldor and Hicks proposed their notions in the 1930s and the 1940s and there is this immediate reaction where a lot of problems were discovered. For example, problems of intransitivity—Scitovsky talks about that.5 Indeed, one of Arrow’s motivations in his book was thinking about Kaldor-Hicks efficiency.6 But at the level of applied economics and in terms of governmental practice, Kaldor-Hicks efficiency is seen as the touchstone. And also, I think—at least in the U.S.—in Law and Economics, where it is associated with the intuitive notion of making the pie bigger. One often hears the expression, ‘just make the pie bigger, why isn’t that a good thing?’. Eric Posner and I were driven to respond to that. Also, part of what motivates Kaldor-Hicks efficiency is the notion that one cannot make interpersonal comparisons. So, Eric and I did a double move. One was to say, ‘of course, we can make interpersonal comparisons’; and the second was to think of cost-benefit as a rough proxy for overall well-being and to argue that justifying cost-benefit analysis by way of overall well-being is much more persuasive—at least if overall well-being has some kind of intrinsic normative weight. Kaldor-Hicks efficiency doesn’t.

---

4 A Kaldor-Hicks improvement, named for Nicholas Kaldor and John Hicks, is an economic reallocation of resources such that those that are made better off could hypothetically compensate those that are made worse off, so that a Pareto improving outcome could be achieved. A situation is said to be Kaldor-Hicks efficient, or equivalently, is said to satisfy the Kaldor-Hicks criterion, if no potential Kaldor-Hicks improvement from that situation exists. See Kaldor (1939) and Hicks (1939).

5 What Tibor Scitovsky demonstrated was this. It is possible that if an allocation \( x \) is deemed to be superior to another allocation \( y \) by the Kaldor-Hicks criterion, then by a repeated application of the same criteria, we can also show that \( y \) is superior to \( x \). See Scitovsky (1941).

6 See Arrow (1951).
Thanks, we’ll come back to this later. But let’s start by zooming in on some of the parts of your trajectory that you described before. Starting from the early days, what made you choose history?
I was lucky to go to Yale college, which is a wonderful place. And one of the great things about the undergraduate major there is that you can major in a subject without having that be your exclusive focus. I actually went in thinking that I’d be pre-med and afterwards go to medical school. Then I was in this wonderful, intensive freshman humanities program called Directed Studies, which is one of the famous things at Yale—in six classes over the course of the freshman year, you read great books in literature, political philosophy, and philosophy. And after that I realized that I really wanted to spend a lot of time taking classes in the humanities. Although I also took a lot of economics, I took a fair bit of math, I took some science. But Yale was very, very strong in the humanities, including history—they had a lot of great professors in history.

Also, in terms of my background, I’m Jewish. My family, grandparents, great-grandparents, left Europe because of antisemitism in the early 20th century; that led me to be interested in studying German history. So, I ended up majoring in history. I then received this scholarship to go to Oxford to get a Master’s in history, which I enjoyed very much. But I realized that I didn’t really want to go on to do a PhD in history—that I had more theoretical interests. I feel lucky that I did history, but also that I had all these classes in math in secondary school, and then in college—along with history—economics, math, philosophy, and so forth, that I could then use when I pivoted to more theoretical pursuits.

Do you remember any specific book or set of books that left a deep impression on you from your undergraduate days?
That’s interesting. In Directed Studies we read a lot of different things. We read The Iliad, for example, and The Odyssey, which are incredible. We read Kant (in translation, of course) and we read Hume, and the funny thing is that at the time I was more of a Kantian. I had a good friend and we had debates: he was a Humean, and now I’ve ended up being much more of a Humean than a Kantian. But I was exposed to both Kant and Hume.

After you did your history studies, you went on to work as a management consultant for a while. How did that come about?
I finished my Master's studies at Oxford in 1986. It was a great time to be there, wonderful place, and the dollar was very strong so one could live well as a student. There were a lot of Americans at Oxford who were there via various scholarships and what have you, but then, when that was coming to an end, we were trying to decide what to do. In 1986 Wall Street was booming and, I mean, that sort of seemed interesting—maybe I should make a little bit of money. So, I worked at Oliver & Wyman—which is now a huge firm, but when I worked there it was quite small—and I enjoyed that, but I realized after two years that I wanted to do something more academic. I was fortunate at that point to be admitted to Yale Law School. I went to law school thinking that if things worked out, I could become a law professor and if not, law seemed pretty interesting and there are obviously a lot of things you can do with a law degree. So, it was a short stint in the private sector. (And, by the way, some of the skills that I developed in the private sector still prove useful. Chapter five in Measuring Social Welfare, where I have this sort of simulation model of risk regulation, is based on an Excel spreadsheet. I learned to do Excel as a management consultant.) In a nutshell, I went to law school with an eye to getting into the legal academy and thinking that law would engage my theoretical interests more than a history PhD would do.

You also served as a law clerk in the U.S. Court of Appeals for Judge Harry Edwards and later in the U.S. Supreme Court for Justice Sandra Day O'Connor. What can you tell us about these experiences? Have they shaped you in any particular way?

They were really wonderful experiences. I was very fortunate to get those jobs. What's interesting is that, as a law clerk you are really working for the judge. For example, each of the nine Supreme Court justices has four clerks. You are working for the justice. You are hired by the justice. The four clerks work in chambers. And it’s a very small unit: there is the justice, the four clerks, and maybe a couple of assistants, but that’s chambers. You also interact with other clerks. The same is true on the Court of Appeals. At that point, maybe there were 10 or 11 judges, each had three clerks. So, it's very intimate, and intense—huge amounts of work.

The main job of the clerks was twofold. First, before the case was decided, you as a clerk would write a memo for the judge analyzing the case. And that was supposed to be your impartial take on it. Neither the judge nor the justice ever said to me, ‘this is the way you should analyze it’. It was rather, ‘here are the briefs, here is the opinion, write a 10- or 15-page
memo analyzing the issues here’. You’d then discuss that with the judge. Then (and this was the second main part of the job) the judge might well end up writing an opinion in the case and you were told to prepare a first draft of the opinion. And the opinion would then be circulated to other chambers. It was a very intellectual experience. Which in a way was kind of inspiring. Of course, judges have political views, and there might well have been some politicking behind the scenes among the judges, but fundamentally that was not what the clerk’s job was about. The job was doing legal analysis.

There were many issues on which the analysis was helpful. That is, after drilling down on the questions in the case you’d realize that one answer was the legally right answer, that one answer was the more persuasive answer. Now, and this is really at the Supreme Court level, there are cases where the court is simply not bound by prior law, where the court in effect is making law—although the court doesn’t say that. But even at the Supreme Court level, and certainly at the Court of Appeals level, there are lots of cases where, again, this exercise of legal analysis could point you to a right answer, a better answer. It’s a good question, which I haven’t really thought about, but I think that the experience as a law clerk of finding right answers through legal research has in turn helped to motivate and inspire me with respect to my scholarship.

Would you say that was a way of going to philosophy graduate school? Because you are reading a lot of texts and there is a pressure to read these texts quickly and analyze them in a very impartial way. That mirrors a lot of graduate training in analytic philosophy at least. That’s interesting. It’s a good question. Being a law student and then law clerk are highly analytic. And actually, the thing I like most about teaching is the analytic part. People say, ‘oh, you teach law school, you teach legal doctrine, that’s boring’. But for me the best thing about the classroom experience of law teaching is that it’s analysis. It’s not memorizing the rules. What it is, is: ‘Okay, so this is the case, this is the issue, should the court decide the issue this way or that way? How will we analyze that using these legal materials?’ That’s analytic reasoning. And the same is true, obviously, as a clerk. The difference between judging and philosophizing is that judges are constrained by the legally authoritative texts. That’s it. You’re never going to say: ‘Well, why do we have this constitution? Maybe we should change the constitution’. You cannot. There’s a limit to how deep you can go. But philosophers are bound by nothing.
I’ve already mentioned this, but some of my research development is just the serendipity of my academic career. I was fortunate to get a teaching job in ‘95 after a brief time in practice. And I just happened to have colleagues who were really good philosophers, I mean, really good philosophers. And so, I experienced arguing with them, about bedrock stuff. For example: we all have these deontological intuitions, ‘you shouldn’t kill someone, you shouldn’t kill someone even to save five lives’. But to really drill down on deontology, and to make sense of its content, scope, and justification, is not easy. The people I was working with at the University of Pennsylvania Law School at the time, people like Michael Moore, Heidi Hurd, and Steven Perry, were some of the greatest deontologists and legal theorists. I ended up not going in the deontological direction, but the depth of their involvement and the seriousness of the analysis were really inspiring to me.

Talking about moving from working as a clerk, to coming back to academia. In a more light-hearted way, in Measuring Social Welfare, you use the following example to drive home the point that preference satisfaction accounts of well-being can be problematic if people’s preferences are based on poor information:

Felix, remembering all the courtroom dramas he has watched, prefers to be a litigator over alternative careers. If he were better informed, he would realize that much of a litigator’s time is spent reading documents or writing briefs—which doesn’t interest Felix—and not making dramatic closing arguments. Felix’s actual preference for the litigator career diverges from what he would prefer, career-wise, if he knew more. (Adler 2019, 48)

You, yourself, worked as an associate at the litigation department of Paul, Weiss, Rifkind, Wharton & Garrison (New York) in 1994. Is this example by any chance autobiographical? That’s funny. Maybe a little bit. Paul Weiss is a great firm, a great litigation firm. Not just commercial practice, but also pro bono practice. For example, much of the litigation against the Trump ‘Muslim ban’ was undertaken by Paul Weiss. So, it’s really a fantastic firm. But yes, it became clear to me pretty quickly that I didn’t want to be a litigator. After my law clerkships, we had moved back to New York from Washington D.C. because my wife was finishing her PhD in history at Columbia. I said to Paul Weiss, you know, I’m likely to go on the law teaching market and try to get an
academic job, I'm probably not going to stay here. And their response was, 'that's fine'.

But the story I like to tell about these large law firms—the term that is often used in the U.S. is 'Biglaw'—is this. In the Biglaw firms, typically the time it takes to become a partner is maybe eight years or ten years. So, the image I had, before I started at Paul Weiss, was: you have this group of associates, and everyone sits around for eight years or ten years, and at the end of that they pick a couple to be partner. That's not true. What happens is that you have people who are associates who, at these big firms, are typically very talented people. And most of them decide that they don't want to become Biglaw partners. That is, people end up going into the academy, going to work for government, going to work for a smaller firm, going to work in-house at a firm, getting out of law entirely. Because what it takes to become a partner at that kind of firm is not just intelligence, but also a tolerance for an incredible amount of hard work. And a willingness to have the work be driven by clients. Because at the end of the day, at that kind of firm, you have a client: you are litigating a matter because the client comes in and says, 'here is this problem I need you to solve'. So, it's those particular skills, motivations, and interests that it takes to become a partner. And so, what in fact happens is that most people say, 'that's great, but I'll stay here for a couple of years to get the training and then I'll leave'. And at the end of the eight or ten years there are only few associates remaining who still want to become partner, and one of them is chosen. Again, I was very happy during my year at Paul Weiss. I think, however, that I would not have been a good litigator. In part because I really value intellectual freedom. I mean, one of the great things about being an academic is control of your own time. I just value the ability to decide that today I'm just going to read this book and think about this. I was very fortunate to get the job at UPenn Law School, I was there for 18 years. That worked out very well and then I moved to Duke.

You already mentioned that there was this idea to pursue a career in legal academia, but the courses you started out teaching were in Administrative Law, Food and Drug Law, and Theories of the Administrative State. This, at least for us, sitting in Europe or India, or wherever else, seems like a world apart from the work you're known for these days, at least among the readers of this journal. What led you to start working on cost-benefit analysis, in particular, and the foundations of
**public policy, more generally? You’ve already spoken about the influence of Eric Posner in this regard. Is there anything more that you want to say about this?**

One of the things I should say is that the U.S. Court of Appeals for the D.C. Circuit is the federal appellate court that deals with a huge volume of administrative law issues. Administrative law in the U.S. is the law governing administrative agencies—for example, the Environmental Protection Agency, the Food and Drug Administration, the Department of Transportation, and so forth. A lot of the cases involving challenges to rules or decisions by those agencies go to the D.C. Circuit. As a law clerk on the D.C. Circuit, you will have a year during which a lot of the work will be on administrative law. You will come out of that either loving administrative law or hating administrative law. For whatever reason, I came out of that clerkship just loving administrative law. The issues are often somewhat technical, they’re quite detailed, but I found that really interesting.

I actually thought that what I wanted to do as a law professor, was to write a treatise on administrative law. I started law teaching with that in mind because, back in the 1950s, ’60s, and ’70s, there were these great books about administrative law—I mean really, really good books about the doctrine. And then, of course, I got involved in other things. But that’s why I started teaching administrative law, because of my experience as a clerk on the D.C. Circuit. And as a new professor your Dean says, ‘you’re going to teach this’. So I was like, ‘yeah, sure, I’ll teach administrative law, teach food and drug law, theories of the administrative state, this is all about administrative law, this is where I start’.

Cost-benefit analysis is very important to U.S. administrative law. Since the Reagan administration, major administrative agencies, by presidential order, have been required to engage in cost-benefit analysis. Basically, what the executive order says is that all regulations, unless the statute requires a different approach, should satisfy a cost-benefit test. And, in addition, for large regulations, agencies have to produce extensive cost-benefit documents that are then reviewed by OIRA (the Office of Information and Regulatory Affairs), a cost-benefit oversight office. So, it’s really because I was teaching administrative law that I got interested in cost-benefit, which in turn led to writing with Eric about that, which in turn led to... It was a sort of winding path. But every step made sense at the time. It’s just that when you look back you may think, ‘how on earth did you end up moving from administrative law and food and drug law, to writing about measuring social welfare?’.
This leads to our next question. Like you said, you have written extensively about cost-benefit analysis and also about the foundations of public policy. So, what field would you call your academic home? Would you say you're doing Law and Economics? Or Philosophy of Economics? Or maybe PPE? Or something else entirely?

In a way I'd say Law, Economics, and Philosophy. Something I have been persuaded by throughout is that there are huge returns to engaging philosophically with the foundations of economics. And one of the things that traditional Law and Economics doesn't do enough, and certainly economists don't do enough, is to engage with the philosophy. So, I think philosophy is going to have to be in there, which is why I wouldn't just say 'Law and Economics'. I wouldn't say philosophy of economics, because that in a way takes economics to be primary. I mean, I think philosophy is equally important. Let me say that my interests are normative. At the end of the day, my aim is to try to figure out what to do, to provide normative guidance. And I think Philosophy, Economics, and Law are on a par with respect to that aim. Philosophy helps us in drilling down on these basic questions, such as, ‘should we be consequentialists? Non-consequentialists? What's welfare?’ And so forth. Economics—welfare economics—is hugely significant in providing all of these formal results and concepts that are very important for my purposes. One could also say that it’s Philosophy, Economics, and Policy. Concerning PPE, I haven’t done much on political theory, so I wouldn’t use that label. So, let’s say Law, Economics, and Philosophy.

That's interesting. You said, your interest is in practical reasoning, ‘what ought we to do?’, either at an individual level or socially. That's the motivating question. And you go to whichever discipline is relevant to answer that question. The thing is, some of these disciplines require a kind of training in formal reasoning, not just to contribute to that literature, but even to understand that literature. But your background was in history and law. Did you always have some 'formal chops' as part of your training, or is it something that you had to learn 'on the job' to do the kind of things you wanted to be doing?

At the end of the day, ethics is about what to do. And there are different relevant tools—there’s conceptual reasoning, there’s the process of arriving at a reflective equilibrium, which is philosophy. I was lucky to have a really, really good secondary school education. I went to a school in New York called the Fieldston School, which is an excellent school. And then I
had a really good undergraduate education at Yale. So, for example, in
college, I took multivariable calculus, I took real analysis, I took mathe-
matical economics, I took game theory. I also, in Directed Studies, did
enough philosophy to have some background in that field. So when I de-
cided that I wanted to be more engaged in philosophy, be more engaged
in economics, I had enough tools.

One of the things I say to students is, ‘take as much math as you can,
take as much philosophy as you can, so then you have the tools’. You can
read these literatures and then you can decide what to research. And, by
the way, if you're interested in something and you feel like it's too hard
for you to make a contribution, you can find a co-author.

This leads nicely to our next question. Throughout your career, you
have co-authored both books and articles with scholars both within
your discipline (law) but also other disciplines. Can you tell us a bit
more about these experiences and the value you see in these types of
projects? And, further, what might be some of the barriers to collabo-
rate in and across disciplines?

I've been lucky to collaborate with interdisciplinary people. So, again, my
first collaborator was Eric Posner, who is probably at this point one of the
most famous law professors. He's been incredibly influential and is in-
credibly interdisciplinary. We were both law professors who were doing
interdisciplinary work. There's Chris Sanchirico, who’s a law professor
but with a strong background in economics. I've worked extensively with
Marc Fleurbaey. We co-edited the Oxford Handbook of Well-being and Pub-
lic Policy, and Marc is a serious economist but one who also has many,
many philosophical interests. And similarly, I've done a lot of work on
applying social welfare functions to risk regulations with a group that
includes Jim Hammitt, Nicolas Treich, and Maddalena Ferranna.

You have to find a co-author that you get along with. That you enjoy
working with. Who brings something to the table that you don’t have (and
the converse also has to be true). But I also think it cannot just be strate-
gic: you cannot say, ‘I don’t know this, I need this, so I'm collaborating
with this person’. The collaboration has to really, really work. The most
important thing I have found is, ‘do the work’. A collaboration where you

---

7 See Adler and Posner (1999, 2000b, 2006, 2008), as well as the co-edited volume Cost-
8 See Adler and Sanchirico (2006).
9 See Adler, Hammitt, and Treich (2014), Adler et al. (2020), Adler et al. (2021), Adler et
al. (2023), and Ferranna, Hammitt, and Adler (2023).
expect that the other collaborators are going to do more than you is not going to work. All my collaborators have been equal partners, they have done huge amounts of work. I’ve seen these multi-author collaborations where one author thinks another author is going to do something, and vice versa, and it falls through. Everyone is going to have to be fully engaged. And there is going to have to be flexibility. You cannot insist that the article is going to have to be written in this way, or that this result is important as opposed to that result.

This recent book, *Prioritarianism in Practice*, which appeared in 2022, involved a whole team of people writing about the application of prioritarianism in different areas.¹⁰ There is absolutely no way I could have done that myself. One of the things I would say about edited volumes more specifically—I have done a bunch of these edited volumes—is that the way to push along an edited volume is to have a conference and ask people to present the first draft of their chapter at the conference.¹¹ Contributors can delay writing a first draft indefinitely over email, but no one likes to show up at a conference empty-handed. I’m grateful to Duke University for funding the Center for Law, Economics, and Philosophy, which I’ve used to organize conferences.

*Is there any way in which collaborating has influenced you?*

Yes. For example, Eric is incredibly successful because whenever he writes, he wants to know what the bottom line is. He’s had a huge influence across many, many different fields. So, for example, now there is this movement in the U.S. about anti-trust with respect to labor relations, breaking down the way in which firms have used monopoly power to limit the opportunities of workers. And Eric has written a seminal book about that.¹² The point is, he’s very focused on, ‘OK, so, what are we saying? What’s the recommendation? Can we say this clearly without losing the forest for the trees?’ So that was very helpful to me. If I had written *New Foundations of Cost-Benefit Analysis* myself, the book would have been three times as long and much less influential.

*Maybe you could say something more about your personal convictions. You said before that at UPenn you had colleagues who had more of a*

---

¹⁰ See Adler and Norheim (2022).
¹¹ Edited volumes include, Adler and Posner (2000a), Adler and Himma (2009), Adler and Fleurbaey (2016), and Adler and Norheim (2022).
deontological bent of mind. But you stuck to a more consequentialist viewpoint. Is that a conviction you always held?

Jules Coleman—a great legal philosopher, whose legal theory class I took at Yale—said something to me once that really stuck in my head, which is: ‘As scholars, we cannot answer the ultimate questions, but we can answer the penultimate questions’. If I write a book taking on the debate between consequentialism and deontology, I’m not going to make much of a contribution. It’s not like I’m going to suddenly persuade deontologist to become consequentialists. That’s just not how philosophy works. But what I can do is make progress within consequentialism. I find welfare-consequentialism persuasive. Is my credence one hundred percent? No. I find it to be persuasive, plausible, and worth pursuing.

Maybe this is something I learned as a management consultant, which is: as a management consultant, we always advise companies to pursue market opportunities. As a scholar, of course, if you write about something nobody cares about, nobody is going to read your work. But if you write about the thing that everybody else is writing about, you’re also not going to make much of a contribution. So, it struck me that there was a lot of work to be done fleshing out welfare-consequentialism. Both at the level of theory, and at the level of application. And I find welfare-consequentialism to be quite plausible. So, that’s what I’ve done.

II. WELFARE-CONSEQUENTIALISM AND SOCIAL WELFARE FUNCTIONS

Let us then turn to welfare-consequentialism, and more specifically the work you have engaged in, in recent years, since the publication of New Foundations of Cost-Benefit Analysis. This work is helpfully summarized in the content and title of your last book, Measuring Social Welfare. Is it right to say that this has been motivated by the need to find a better alternative to cost-benefit analysis as a basis for public policy evaluation? If this is correct, which you also hinted at in your previous answers, why do you take the view that cost-benefit analysis is an insufficient basis for public policy evaluation? And how have your views on this evolved over time?

Indeed, a lot of my work, starting in 2006, has been motivated by the need to improve on cost-benefit analysis through this notion of the social welfare function (SWF). I see myself as proceeding on three tracks. The first is filling in the philosophical details—for example, the nature of well-being, or the functional form of the social welfare function. The second is
applying the social welfare function in different policy domains, showing that this framework is actually useable. There is, of course, this literature in economics where SWFs are used: optimal tax theory. James Mirrlees wins the Nobel Prize for using social welfare functions to think about taxation. And, also, climate economics. But there are other fields where it’s not used much. My view is that the social welfare function framework is an all-purpose policy tool. It’s important to demonstrate that—to persuade people that it’s actually useable, that it’s not just a theoretical device. So, the second aspect of my work has been to do that. I’ve written about social welfare functions and both risk regulation and, to a lesser extent, climate economics. And this Prioritarianism in Practice project was an attempt to have lots of people in different policy areas doing that. And then the third track has been to make the apparatus more accessible. Measuring Social Welfare was meant to provide an overview of social welfare functions that was sufficiently rigorous but more broadly accessible than Well-Being and Fair Distribution. As I already said, all three tracks are intended to get us to move beyond cost-benefit analysis—with cost-benefit analysis as the target because this is the dominant policy analysis approach in the U.S., where I live and teach, and in applied economics.

It seems to me that there are really three justifications for cost-benefit. There is the so-called neoclassical justification, which points to Kaldor-Hicks efficiency. That’s problematic for the reasons I have alluded to, which go back to Amartya Sen. Either the policy is actually Pareto superior, in which case just say that; or, merely potentially Pareto superior, in which case there is a potential—an unrealized potential—to convert the policy into an actual Pareto improvement, which doesn’t seem to be much of a justification. The second defense of cost-benefit is the defense that Eric and I mounted in our work: cost-benefit analysis is a rough proxy for overall well-being. But why have a rough proxy? Can we not do better? Cost-benefit analysis uses money as the measuring rod for well-being. But money is an imperfect measuring rod given diminishing marginal utility.

---

13 James Mirrlees was, together with William Vickrey, awarded the 1996 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (see Royal Swedish Academy of Sciences, n.d.). Mirrlees’ work—building on the work of Frank Ramsey (1927)—laid the foundations of optimal tax theory (see Mirrlees 1971).
14 See Botzen and van den Bergh (2014) for an overview of the use of SWFs in climate economics.
15 On the application of the SWF approach to risk regulation, see Adler, Hammitt, and Treich (2014), Adler (2015), Adler et al. (2021), and Adler (2023); on the application of the SWF approach to climate change, see Adler and Treich (2015), Adler et al. (2017), and Adler (2018b).
The third justification for cost-benefit, which I think is probably the best, is the Kaplow notion. Kaplow actually accepts a social welfare function as the foundation for policy analysis but says that we should limit the direct use of social welfare functions to the tax system, and we should use cost-benefit analysis for non-tax policies.\(^\text{17}\) I think that’s not fully persuasive for various reasons, but again note that Kaplow and I agree that the foundation in terms of designing government and in terms of figuring out what policy techniques different institutions should use, is a social welfare function and not cost-benefit analysis.

**So, to label stages as well. The social welfare function approach that you propose has three key components. First, there is a well-being measure, which translates each of the possible outcomes (a complete description of a world) of a policy choice into a list (‘vector’) of interpersonally comparable well-being numbers, which quantify how well-off each person in the fixed population would be in that outcome. Second, there is a rule for ranking alternative vectors, or lists of well-being numbers. Third, there is an uncertainty module, which orders policies understood as probability distributions across outcomes. So, a few questions of clarification about the framework, before we get to more substantive issues about this framework.**

*First, there are further sub-components to this approach which consist in the restrictions imposed on the three basic components (and we will get to some of these restrictions below). This has the implication of ruling out a broader class of social welfare functions. Those that do not allow interpersonal comparisons of well-being, for example; or, non-welfarist social welfare functions, for example. What was the motivation for introducing and scrutinizing only those social welfare functions that conform to the constraints imposed on the distinct components? We ask because the social welfare function approach is more general than the class of social welfare functions discussed in the book.*

The social welfare function approach, as I see it, is based on the social choice, or the theoretical welfare economics, literature on so-called social welfare functionals—the SWFL approach—which originates with Amartya Sen.\(^\text{18}\) There are various good review chapters, but, for example, Bossert and Weymark have a fantastic chapter.\(^\text{19}\) The SWFL approach imagines

---

that we have a set of outcomes and a profile of utility functions for \( N \) individuals, and then we have a mapping from a given profile of utility functions to a ranking of the outcome set. The so-called welfarism theorem shows that if you adopt certain assumptions—in particular, the independence-of-irrelevant-utilities axiom—this mapping takes the form of a single ranking of utility vectors. The SWFL approach is somewhat narrower than the approach I adopt, because of the assumption that we should understand people’s well-being in terms of utility. Each person has a utility function mapping the outcomes to utility numbers—a utility function that tracks her preferences.

What I do is that, instead of having a profile of utility functions, I have a well-being measure. This well-being measure converts an outcome into a vector of well-being numbers. In that sense it’s more general than the SWFL approach. I am adopting the welfarism assumption. The most general formulation of my approach would be that we have a well-being measure, a set of outcomes, and a function that takes any given well-being measure to a ranking of a set of outcomes. But I’m assuming, as per the welfarism theorem, that that mapping will take the form of a single ranking of well-being vectors.

Now, one of the things that the SWFL literature does is to look at different possibilities for this ranking, including possibilities that don’t make assumptions about interpersonal comparisons. For example, various kinds of dictatorships. A result that I cite in chapter two of *Measuring Social Welfare* is that, if well-being is not interpersonally comparable, we will end up with a dictatorship. In the book, I have this notion of ‘admissible rescalings of the well-being measure’, which is based upon the SWFL literature. Suppose that we have a theory of well-being, which could be a preference-based theory, a hedonic theory, an objective-good theory, or some other kind of well-being theory. That theory is going to tell us what the admissible well-being comparisons are. In particular, it’s going to tell us if we can make, first, intrapersonal comparisons of well-being levels, differences, and ratios, and, second, whether we can make not just intrapersonal comparisons of that sort but also interpersonal comparisons. Corresponding to this specification of admissible well-being comparisons is going to be a set of admissible transformations of the well-being measure. So, in particular then, if we believe that the only admissible well-being comparisons are intrapersonal comparisons of levels and differences, then any individual-specific positive affine transformation of the well-being measure will be admissible. An individual-specific positive affine
transformation is going to disrupt the interpersonal information; but if you think that such interpersonal information doesn’t exist, this transformation is admissible because it will preserve all the intrapersonal information about well-being levels and differences. But then there is this theorem, which says that, if we require the social welfare function to be invariant to individual-specific positive affine transformations, and also satisfy the Strong Pareto principle, we are going to end up with a serial dictatorship. Non-dictatorship is a weaker requirement than anonymity, which in turn I take to be fundamental to ethical reasoning. In short, if I want a social welfare function to satisfy at the very minimum non-dictatorship and more broadly anonymity, I’m going to require some interpersonal comparability.

*The second question of clarification is, again, why the identification of the social welfare function approach with well-being. The object of inquiry in the book is well-being, but the social welfare function approach can also be deployed in the evaluation of other distinct moral concepts, like justice. Justice might be related to well-being, but it’s not the same thing. We can also use the social welfare function approach to measure poverty, to measure inequality. What was the motivation to identify the social welfare function approach as an approach to evaluate well-being?* 

Richard Arneson, who is a great philosopher, has an article, “Welfare Should Be the Currency of Justice”. I don't view a focus on well-being as being inconsistent with a focus on justice. There is a literature, going back to work by Jerry Cohen, Sen, Dworkin, Arneson, that asks: ‘To the extent we are concerned about distribution’—and I’m concerned about distribution—‘what should the currency be?’ Rawls says that the currency for distributive justice is primary goods. Rawls talks about the principles of justice applying to primary goods. My view is that, to a first approximation, the currency for justice is well-being. But we can then think about what a just allocation of well-being would be. For example, we can say that the correct view regarding the just allocation of well-being is not utilitarianism, it’s, let’s say, egalitarianism, or leximin, or prioritarianism.

My view is that a concern for justice, or a fair distribution, and consequentialism are not mutually inconsistent. And that a concern for justice, or a fair distribution, and a concern about well-being are not inconsistent.

---

20 See Adler (2019, 45–46).
21 See Arneson (2000).
23 See Rawls (1971).
Indeed, my 2012 book is titled Well-Being and Fair Distribution. What I was trying to get at with that title was that you can be concerned about fair distribution and about well-being. And the same with both poverty and inequality. In practice, poverty metrics and inequality metrics are often applied to income. But if you view income there as just a proxy for well-being, the view is going to be that, if you care about poverty and inequality, it would be better to apply the metrics to well-being.

A last question of clarification. What was the motivation to apply the social welfare function approach only to public policy? The field of Law and Economics has a long tradition of applying tools from economics to a variety of domains, like contract law for example. (Like the Coase theorem.) Do you think an avenue for future research is to apply the social welfare function approach to other domains of law as well?

Absolutely. Again, this relates to the path-dependence of an academic career. So, again, I graduate law school; I work as a clerk for a judge (Judge Edwards, who is, by the way, a very great judge, he’s still on the bench) where a lot of the work is administrative law; I start teaching administrative law (administrative law is public law, it’s not private law, I have never taught a private law class); I become interested in the use of cost-benefit analysis as part of public law (that is, thinking about big regulations and so forth); so that leads to my interest in social welfare functions and public law. But there is no question that we can think about legal doctrines for an area of law—such as the doctrines of contract law, the doctrines of tort law, or the doctrines of property law—as objects that can be evaluated using a social welfare function. Would it maximize, for example, prioritarian social welfare, if, for example, the doctrines of tort law looked this way as opposed to that way? So, yes, it’s a fruitful question and something, I think, that should be pursued.

Thank you. Young scholars reading: PhDs waiting to be written here! Stephen Hawking is supposed to have said: ‘Someone told me that each equation I included in the book would halve the sales. I therefore resolved not to have any equations at all’. This is relevant here because one needs a certain competence with formal reasoning to engage with the social welfare function approach. Not least because informal reasoning can be quite treacherous in this domain of inquiry. At the same time Hawking is not wrong to suggest that the use of notation can come
in the way of the broader uptake of arguments that use formal reasoning. What are your views on the trade-offs involved in using formal reasoning, especially within domains where it is relevant to public policy? I really think it’s a question of audience. Obviously, if you're writing for a public, even an educated public, don’t use equations. Because even educated people may not be comfortable with equations. But the same will be true of philosophical language. If you’re talking about public policy, don’t use the term ‘agent-relative side-constraint.’ Members of the public don’t know what an agent-relative side-constraint is. Don’t use the term ‘supererogation’. So, let’s be clear: if you're writing a popular book, and I think that's a very important thing, even a book for an educated public, you are going to have to be very careful about using formal notation or about using technical nomenclature.

On the other hand, if you’re writing for scholars, then you should use whatever linguistics tools and approaches, mathematical tools and approaches, match the inquiry. And here I would say, it’s a trade-off. There are people who have very advanced mathematics and are very sophisticated as a conceptual matter. That’s really difficult to do. So, obviously, at the high-end, someone like Amartya Sen. But most of us are not like that. So, there is a trade-off. The level of mathematics that you would use in a paper in *Econometrica* is going to be higher than the level of mathematics that you’re going to use in, say, *Economics & Philosophy. Economics & Philosophy*—which I’ve been privileged to edit for a number of years—does allow math, but we are not going to have the level of math that you might have in the most mathematical journals. On the other hand, the level of conceptualizing that would be tolerated, or demanded even, in a philosophy journal such as *Ethics* is not going to be acceptable in an economics journal. So, there is a trade-off. But I would say that, if we are talking about welfare economics, normative economics, and its twin in philosophy, which is moral philosophy, scholars in those fields should have some degree of comfort and openness to both conceptualizing and mathematics. Because it seems to me that both of those tools have been incredibly fruitful. And scholars who say, ‘I'm just not going to do it—if there is math there, I'm not going to read it’, or, conversely, ‘if there is conceptualizing, I'm not going to engage in it’, I think, are just making a mistake.

Now, the third question is, what if you’re writing not just for other scholars, or for the public, but for students? That’s in a way an exercise in textbook writing. In writing *Measuring Social Welfare*, I felt like I was
writing for students, and also, by the way, for law professors. You may have colleagues in other fields who are not deeply engaged in either economics or philosophy, but might be interested in those topics, and so, at that point, you try to lower the barriers of entry. But at the end of the day, it seems to me that a lot of the power of the social welfare function approach is the mathematics. At two levels: one is that a lot of the results, the most important results, depend on the mathematics—so, Arrow’s theorem, or Harsanyi’s aggregation theorem, for example, depend upon the mathematics. The other thing is that the tool itself is a mathematical tool—this is the point. The tool allows for systematic and tractable policy analysis by converting outcomes into vectors of well-being numbers; by converting the problem of policy situations under uncertainty into the problem of maximizing the expected value of some ethical utility function; and so forth. So to say, ‘let’s just drop mathematics’, is just to say that we are not going to use the tool as it stands.

Now, to go a little deeper into the weeds. The social welfare function approach that you present is consequentialist, in the following sense: a policy $P$ ought to be pursued only if the state of affairs $x$ resulting from $P$ is ‘optimal’ or ‘best’ with respect to a given rule. By that we mean that it’s at least as good as any other outcome or state of affairs that could have been pursued instead. Now, some deontologists might take issue with the whole set-up—and that’s as you said earlier unproductive as a debate to engage with—but our question here is internal to consequentialism: Why think that consequentialism requires optimization? Alternatively, you might respond by weakening the consequentialist claim to the following: a policy $P$ ought to be pursued only if the state of affairs $x$ resulting from $P$ is ‘maximal’, by which we mean it’s not strictly worse than any other state of affairs $y$ that could have been pursued instead (with respect to a given rule). Then, our question is why wouldn’t consequentialism require optimization or going for the best? In short, what sort of action-guidance is required, you think, of consequentialism?

That’s interesting. I view consequentialism as the view which says that ethical guidance is ultimately grounded in a single—that is, an agent-neutral—ranking of possible worlds. And specifically, I’m a welfare-consequentialist to a first approximation. So, ethical guidance is grounded in an agent-neutral ranking of possible worlds that satisfies certain basic
axioms, in particular Pareto-indifference, strong Pareto, and anonymity, that express the core of welfarism.

By the way, what does the ranking involve? I believe that it’s a well-behaved ranking, so I believe in transitivity, but I believe that the ranking takes the form of a quasordering, or a preorder, which allows for incompleteness—there could well be incompleteness in the ranking. Then we have the question of ‘what to do?’ What action should we choose in light of the ranking? In terms of my work, that’s what I call the ‘uncertainty module’.

There are lots of further assumptions one might make to make consequentialism tractable, but that I don’t think are essential to the approach. Number one is assuming that well-being is measurable. Actually, in *Well-being and Fair Distribution*, I allow for the fact that there are going to be pockets of incomparability in terms of well-being itself. Let me put it this way: we have an agent-neutral ranking of worlds that’s grounded in well-being. Well-being itself, or the well-being account itself, is captured in a ranking of what I call ‘histories’, where histories are pairings of individuals and outcomes—such as, ‘being Matt in this world’. And also a ranking of pairs of histories, so as to capture differences. Those rankings too—the rankings of histories and history pairs—are quasiorders. So, there can be incompleteness at this foundational level with respect to well-being. The assumptions that the well-being rankings are complete and measurable by a well-being measure, and that the social welfare function takes the form of a complete ranking of well-being vectors, I mean, these are incredibly important for tractability, but I don’t view them as being part of the bedrock of consequentialism.

Now, finally coming in a long-winded way to your question: there is this further question of ‘what should we do?’ Do we always optimize? Some people believe we should satisfice. I actually think the most important thing is to have the rankings. We want to go from this ranking of outcomes to a ranking of choices. And if there is a best choice, it seems to me—I’m not a satisficer—we should choose the best choice. But the most important thing is to have the ranking. And, by the way, if the set of choices is infinite and not compact, there may not be a best choice—in that case it’s possible to have an infinite sequence of choices, each better than the one before. It seems to me that fundamentally what consequentialism is about is arriving at a ranking of choices, which might have incomparabilities, from this welfare-based ranking of outcomes, which also might have incomparabilities.
So, the idea is that there is a ranking. The ranking is what constitutes the basis of the choice. Other than that you have no fixed views on action guidance: if you want to optimize (and if you can optimize), do it; if you want to maximize, maximize; if you want to satisfy, satisfy (and so on)?

This is very useful for me, because it summons me to be very precise as to what my views are. It seems to me, the most important thing is to develop this welfarist ranking of worlds and to use that to arrive at a ranking of choices, for any choice situation that an agent might confront. That’s the guidance. The guidance is the ranking. And then on top of that, it seems to me that, if given the ranking and the choice situation there is a best choice, then you should choose it. But there might not be. And given incomparabilities it may be that the best you do is to maximize as opposed to optimize.

The previous question is of course whether or not consequentialism requires completeness, as you brought up.

I think not. Joseph Raz’s work has pointed to allowing incompleteness.24

A related question—that has implications for some of the avenues for future research that you specify in the last chapter of Measuring Social Welfare—is this: Does consequentialism require transitivity?

I have always worked with an assumption of transitivity. John Broome’s view is that transitivity just falls out of the logic of betterness: if you really understand what it means to be better or worse, that means transitivity.25 In effect, a conceptual argument. I’m not sure I buy that. It seems to me that asymmetry is part of the logic of betterness. If I say, ‘this cup of coffee is better than this glass of water, but this glass of water is also better than this cup of coffee’, I don’t understand betterness. And, similarly, it seems to me that symmetry is part of the logic of equal-goodness. If I say, ‘this cup of coffee is equally good as this glass of water, but, by the way, this glass of water is not equally good as this cup of coffee’, then I don’t understand equal goodness. I’m not sure I’d say the same thing about transitivity.

---

25 Broome (2004) holds that “transitivity is a feature of the logic of ‘more … than’” (50). Indeed, “that comparative relations are transitive is self-evidence” and, consequently, “it does not itself need to be supported by argument, and not much argument is available to support it directly” (51; see also 1991, 11–12).
On the other hand—and this relates to the point that I see ethics as giving guidance—transitivity is just incredibly useful in terms of its inferential properties. Imagine that I have this cup of coffee, this glass of water, and this bowl of ice cream; if I know the cup of coffee is better than the glass of water, and the glass of water is better than the bowl of ice cream, I can then conclude given transitivity that the cup of coffee has got to be better than the bowl of ice cream. I don’t need to make a direct comparison. If you give up transitivity, then for every pair of items, to figure out whether one is better-than or equally-good-as the other, you need to make a direct comparison of the two, which is just cognitively overwhelming.

I’ve have never adopted a view that ethical criteria are simply criteria of rightness. It seems to me that the main reason, or a major reason, we care about the betterness ranking of worlds is because we want to use it to give guidance to us. And, by the way, we are cognitively limited, so something that is very useful is to have the guidance take the form of a transitive ranking.

The other point is that, some of this is simply the economist’s approach of looking under the streetlight because that’s where the light is: there are just a huge number of results that follow from transitivity. We are just not able to say very much without transitivity. I guess if there were overwhelming intuitively persuasive examples that involve intransitivity, then I might be motivated to give up transitivity. But because I’m not aware of such examples, and I think the deliberational—that is, the tractability—benefits of transitivity are massive, I’m going to, in line with most researchers (Larry Temkin has this research agenda to think about consequentialism giving up transitivity), stick with transitivity.26

III. WELFARISM AND INTERPERSONAL COMPARISONS

The social welfare approach that you present is not just consequentialist but also welfarist. This basically means that it considers only vectors (or lists) of utility numbers that persons realize under different policies, or institutions, or systems. A prominent concern articulated by authors like Amartya Sen and John Roemer is that how that vector was gener-

26 See Temkin (2012).
ated is of no independent moral concern under welfarism, but it is (intuitively at least) of import in assessing well-being. How would you respond to that general criticism of welfarism?

All I would say here is that, in both *Well-Being* and *Fair Distribution* and in *Measuring Social Welfare*, the last chapters in those books have said, ‘until this point in the book, I’ve assumed welfarism, but actually we should move beyond welfarism by taking account of responsibility, by doing what John Roemer has done’. 27 Richard Arneson has been doing this for a long time, in fact Roemer was inspired by Arneson’s work. Arneson says that the currency of justice is strictly speaking not welfare, but opportunity for welfare. 28 I think that’s right. I think the social welfare function approach should be refined to take account of opportunity for welfare in a way that Roemer has done. But there are also a lot of difficult questions that arise in trying to do this.

Would it be correct to say that you see, for example, Sen’s work or Roemer’s work as within the broad class of the approach you’re advocating? You don’t see it as antagonistic to the social welfare function approach—it’s a friendly refinement.

On the question at hand, Sen has written about the possibility of a non-welfarist consequentialism, for example, a consequentialism of rights. 29 That I view as being pretty removed from what I’m doing. I think that the next refinement for the social welfare function approach is to do what Roemer has done, which is to focus not on well-being, but well-being adjusted for considerations of responsibility, opportunity, or desert. That’s motivated by specific intuitions about justice, that there’s a difference between people who end up badly off through brute luck as opposed to option luck. As for a consequentialism of rights, at that point I would object, ‘why should I care about rights as such, I care about welfare’; moreover, the distinction between a consequentialism of rights and deontology, although still there, becomes much thinner.

While the social welfare function approach is compatible with many different views of well-being, like experientialist or objective-list accounts, the view that you focus on in your work is, what can be called,

---

27 See Roemer (1996, chap. 8). For more on Roemer, see his interview in *EJPE* (Roemer 2020).
a ‘laundered’ preference-based view of well-being. Would it be fair to say that that is your preferred conception of well-being, either in terms of what constitutes well-being, or—more in line with what Hausman would call the ‘evidential view’—that suitably laundered preferences are reliable indicators of people’s well-being even though they do not necessarily constitute well-being?

My view on this has actually evolved. First, let me say, there is an ambiguity in the literature on well-being as between preferences and desires. Desires are pro-attitudes towards state of affairs—‘I desire something’. The desires are then either wholly satisfied or not. So, ‘I want a vanilla ice cream cone at some point in time’ and then, later, that desire is satisfied or not. As opposed to preferences, which are what Hausman calls ‘total comparative evaluations’. It’s a total evaluation of, in principle, whole possible worlds or whole lives. Of course, humans cannot think in terms of whole possible worlds or life histories, but at least there can be total comparative evaluations of models of those. That, of course, is the conception of preferences that economists use. So, economists think of preferences as being rankings of whole worlds, or lifetime bundles, or something like that. And then there are utility functions that representing preferences thus understood. So that’s number one.

Number two is that I don’t think of preference-satisfaction as either constituting or necessarily being a good indicator of well-being. There are some people, by the way, who think of desire-satisfaction as constituting well-being, which to me seems to be crazy. Obviously getting what you want is a good thing, but there are lots of other good things. Nor do I think of preferences or preference-satisfaction as constituting well-being. But I also don’t agree with Hausman that they are just good evidence of well-being. I would take a third approach, which is that, to the extent

---

30 The notion of ‘laundering’ preferences was coined by Goodin (1986). In broad strokes, the core idea behind preference laundering is that only preferences that satisfy certain conditions should ‘count’ in (social) welfare analyses. For example, preferences that are formed in light of full information, are based on rational deliberation, and are self-interested. On reasons for laundering preferences and conditions on welfare-relevant preferences, see Adler and Posner (1999; 2000b; 2006, chaps. 2, 5) and Adler (2012, chap. 3; 2019, chap 2).

31 See Hausman (2012, chap. 8).

32 See Hausman (2012, chap 1). According to Hausman, preferences (as they are for the most part, and should be, understood in economics) are judgments (thereby ‘evaluative’) ranking alternatives (thereby ‘comparative’) in terms of everything the agent regards as relevant (thereby ‘total’).

33 See chapter 7 in Hausman (2012) on why the satisfaction of (even suitably laundered) preferences does not constitute well-being, and chapter 8 on why (suitably laundered) preferences can yet be considered reliable indicators of well-being.
that individuals are autonomous persons, we should defer to their preferences with respect to their well-being. So, in deciding whether one life-history or another would be good for someone's well-being, if that person is an autonomous individual, then up to some limit we should defer to her in deciding that. That’s my view—that’s the view that I have come to.

I teach at a law school, and legal scholars think a lot about deference—the legal system is full of deference. For example, officials defer to courts about what the law means. And I think there is a similar structure with respect to well-being. We defer to people who can have preferences with respect to their well-being. Now, the tension is of course that people’s actual preferences might be quite uninformed, irrational, formally problematic. So, I would look to people’s fully-informed—or at least well-informed—and rational—in the sense of meeting certain rationality conditions—preferences. But, of course, there is a tension here with regards to the laundering. For me at least, the rationale for looking to someone’s preferences is that she is authoritative with respect to her well-being, so it’s tied to her autonomy: if we launder them too much, then the preferences have nothing to do with her; if we don't launder them at all, then her autonomy is not fully engaged. Again, my views on this have evolved over time: we want a suitable degree of laundering to capture the preferences she would have were she thinking, or deliberating, autonomously.

That’s interesting. Is it then really about autonomy and less so about well-being? So, would you classify your view as an account of well-being, or is it about respecting autonomy?

No, it’s an account of well-being. For example, I as a moral decision-maker care about someone's well-being. That is, I care about what is good for her, what is good for her life. But I recognize that the person is an autonomous person. She’s autonomous, so I defer to her about her well-being. It’s complicated, it’s not autonomy as an input to well-being, nor is it a concern for autonomy apart from well-being, it's rather that I, in light of the fact that she is autonomous, am deferring to her about her well-being.

This line of thinking also—and I haven’t written about this—helps resolve the question, ‘what is a self-interested preference?’’. This is a huge issue in the literature.34 Preferences have to be restricted in some way. Satisfying my purely ethical preferences doesn’t advance my well-being. But it is not clear how we restrict the preferences. One way to do that is

34 See Adler (2012, 174–181) for a discussion of the importance of, and the difficult in, specifying ‘self-interested’ preferences when making welfare judgements.
to say that the restricted preferences are preferences in light of the person’s judgements about well-being. The moral evaluator is trying to decide whether one life-history or another advances someone’s well-being, the subject there also has a conception of well-being, and so the evaluator is going to defer to the subject’s preferences that are motivated, or produced, by her judgements about well-being.

**So, autonomy is the guiding criterion for thinking about what conditions we should impose on the preferences we use when we do welfare analysis?**

Yes.

**In the past you have talked about conditions of non-remoteness and being fully informed, fully rational.** Are these conditions you still subscribe to?

Yes. Full information and rationality because we want to engage a person’s autonomy, we want the deliberation to be autonomous, which means thinking with good information, thinking rationally. Non-remoteness because there needs to be some self-interest restriction or screen. Some of my preferences are for things that are remote to me in the sense that satisfying those preferences is not going to advance my well-being. This goes back to Derek Parfit’s so-called *Stranger on the Train* example, which at this point is pretty well-known in the literature on well-being.

**In the broader literature on laundering preferences, there is on the one hand a discussion of conditions imposed on laundered preferences for reasons of well-being—full information, non-remoteness, rationality, as we have talked about—but then there are also authors in the literature who discuss imposing additional constraints based on other, non-welfare-related, reasons. Harsanyi, for example, writes that we shouldn’t count people’s anti-social preferences when we do social welfare analysis because that would clash with the reason we are doing social wel-

---

35 See, for example, Adler and Posner (2006, chap. 2) and Adler (2012, chap. 3).

36 The example goes as follows: “Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. Much later, when I have forgotten our meeting, the stranger is cured” (Parfit 1984, 494). Many would agree with Parfit that it seems implausible that the satisfaction of this preference (for the stranger’s betterment) would improve my well-being.
fare analysis in the first place—“general goodwill and human sympathy”;\(^{37}\) Dworkin is worried about double-counting—we shouldn’t use people’s external preferences because then people’s fundamental right to be treated as an equal might be disregarded in the social welfare analysis.\(^{38}\) Are these types of non-welfare conditions something that you would apply to a plausible account of laundered or autonomous preferences?

It’s a good question and I do not necessarily have fully formulated views. First, let us take the example of sadistic or anti-social preferences. I think that, if someone is sufficiently depraved in their preferences, then at a certain point we do not treat them as autonomous. We don’t defer to their preferences at all. We just fall back on whatever our view of well-being is. The picture here is: I as a deliberator have my own view of well-being, which can be a hedonic view, it can be an objective view that mixes various goods—that would be, for example, the view that I would use in thinking about the well-being of children. If an adult is sufficiently depraved, then I am, in effect, going to treat him as a child and not look to his preferences at all. The question then is, ‘what if we have someone who is an intermediate case?’ We are still deferring to their preferences, but some of their preferences are anti-social, or sadistic, or what have you. Presumably, or hopefully, those preferences will be laundered out. If she is not depraved, then with good information and thinking well, those will not be the preferences that she has after idealization. She likes to torture cats, but this is an isolated depravity, so it is to be laundered out.

On Dworkin’s concern about external preferences and double-counting: I would view that as addressed by the non-remoteness condition. My purely ethical preference is not part of my well-being, so that should be screened out by non-remoteness. Now it is complicated, because it is true that, to the extent that I make the realization of certain ethical preferences my own project, those can advance my well-being. Say, I devote my life to protecting a certain eco-system: there is some eco-system, let us say, in Alaska that is endangered, I devote my life to protecting that, so I have not merely an ethical preference that the ecosystem be saved, but it has become my project and that advances my well-being. This again goes

\(^{37}\) See Harsanyi (1977, 647). Anti-social preferences are, for example, preferences based on “sadism, envy, resentment, and malice” (647).

\(^{38}\) See Dworkin (1978). ‘External’ preferences are preferences “for the assignment of goods and opportunities to others”, which can be contrasted with ‘personal’ preferences that are preferences “for [the person’s] own enjoyment of some goods or opportunities” (234).
to how we distinguish between remote and non-remote preferences. Once we have done that, there is not an additional double-counting problem.

There is also the question how you synthesize the preferences of people with different preferences, or formulate what you do with utility functions representing different preferences. Because utility is only unique up to a linear transformation, these different utility functions have got to be scaled in some way. One way to scale them is through a high-low rule. So, there is an option that, for example, everyone strongly dis-prefers, and an option that everyone strongly prefers, and you scale all the utility functions so that you give zero to the first option, and one to the second—one kind of high-low rule. If you think about things that way, there is no double-counting problem. Everyone gets the whole numerical range between zero and one to assign to all their possible histories, from the worst to the very best. The bottom line is that, by laundering preferences, screening for remote preferences, and appropriately scaling utility functions, I think you can handle both Harsanyi’s and Dworkin’s worries.

**Following up a bit more. You wouldn’t impose any substantive conditions on laundered preferences? We might have a worry about idiosyncratic preferences, such as Rawls’ case of the Grass-counter.**

This case takes us back to the relation between objective-good and preference accounts of welfare. On my current view, it is not that the preference account is a competitor to an objective account as a substantive matter; it is rather that we are deferring to the individual with respect to what the balance of objective goods is. Let me also say that, in a way, a hedonic account, which says that well-being reduces to pain and pleasure, or happiness, is simply one type of objective-good account. The hedonic account is an account that says: the only well-being good is pain, pleasure, happiness. At the end of the day, substantively, I would say, any well-being account is going to be an objective-good account of some kind. And then the question is going to be, ‘is the subject sufficiently autonomous that we are going to defer to her about the balance of good, or not?’. If she really just counts grass-counting as the only good, maybe that suggests that she is so depraved that we are not going to defer, but if we are deferring, I wouldn’t build in any substantive conditions.

---

39 The case goes as follows: “Imagine someone whose only pleasure is to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns. He is otherwise intelligent and actually possesses unusual skills, since he manages to survive by solving difficult mathematical problems for a fee” (Rawls 1971, 379).
A final question concerning conditions on laundered preferences: in your 2012 book, you talk about substantive conditions and procedural conditions, as we have talked about, but then also historical conditions.\(^{40}\) One historical condition could be that people’s preferences are not problematically adapted to their circumstances—for example, they are not based on manipulation or on problematic kinds of social pressure. That ties in interestingly to the idea of autonomy. In light of this, how would you look at the problem of adaptive preferences?

That’s a good question. I have certainly taken a stab at the nature of well-being, particularly in chapter three of the 2012 book, but most of my work has not been about that—most of my work has been either on the implications of a preference view for the measurement of well-being, for example, that is my work on extended preferences in the 2016 handbook; or, on the social welfare function framework more generally, which can incorporate a whole range of theories of well-being.\(^{41}\) So, I cannot hold myself out as having well-developed views on this and how to deal with adaptive preferences.

That said, my current thinking would be that we are trying to screen out individuals who are below the threshold of autonomy, or if above the threshold, preferences that are not the result of engaging their autonomy. So that is how we should think about screening out adaptive preferences. And I think that is going to be a matter of historical conditions. It is also going to be information. This goes back to something that Richard Brandt suggests. He talks about cognitive psychotherapy as the process for laundering processes.\(^{42}\) So I would give people information not just about what might happen if their preferences are satisfied but also information about their own history. Part of full information is information about the origin of your preferences. That is going to be part of the solution to adaptive preferences, but it may not be the full solution. Imagine that, through a terrible upbringing, you have been brainwashed into terrible preferences; you are now given information about that, but the brainwashing is such that the information does not dislodge the preferences. But, in general, we can make some progress on the problem of adaptive preferences by conceiving of information about the origin of your preferences—that is, about your own history—as part of what full information means.

\(^{40}\) See Adler (2012, 171).

\(^{41}\) See Adler (2016b).

\(^{42}\) See Brandt (1979), who names the “process of confronting desires with relevant information, by repeatedly representing it, in an ideally vivid way, and at an appropriate time” ‘cognitive psychotherapy’ (113).
This is interesting if we look at moving from laundering preferences in theory—what the conditions would look like and how to characterize laundered preferences—to actually trying to elicit laundered preferences and use them in practice when we do welfare analysis. Presumably there are two ways to approach that: on the one hand, we try to construct some circumstance or a context, when eliciting the preferences, that roughly satisfy whatever conditions we apply, to make sure that people are autonomous—we have some methodological procedure to make sure that people are fully informed about their history and their future and reason rationally; or we make some guess about what those preferences might look like and then impose them on people in some sense. Have you thought about this stage of preference laundering?

That’s a good question, and I'm not an expert on preference elicitation, but I think you are right. For my project, I would be trying to elicit people's well-informed, rational, and restricted (that is, non-remote) preferences. How do I do that? The direct approach is to give them all the information I think they would need and to de-bias them so as to make sure that the preferences meet rationality conditions, such as being transitive and so forth. And then just elicit the preferences. This is a kind of intensive survey method for eliciting these idealized preferences.

Let’s imagine that out of this survey, or based on people's behavior, I get some data and I have to figure out what the preferences are. That is going to be the estimation stage. Presumably, at the stage of estimation, I am going to have to make certain assumptions about what the preferences might plausibly be. Imagine someone that has a preference for eating a saucer of mud but only on Tuesdays—this is an example inspired by Anscombe and Parfit. And imagine that this is a preference that is fully autonomous, that the individual holds with good information and thinking rationally. To me as the elicitor, as the experimenter, that seems so weird—one, a preference for a saucer of mud; two, the preference is only for eating mud on Tuesdays—that in estimating the preferences, I am just probably going to rule that out. At the estimation stage we have to have some model of what the preferences might be. There is an infinity of possible models and I have to rule out some, so my judgement as to which preferences are plausible is going to come into play there.

---

43 See Anscombe (1957) on the example of the saucer of mud, and Parfit (1984), who discusses the case of a man who “always prefers the prospect of great suffering on a Tuesday to the mildest pain on any other day. [...] Simply because the agony will be on a Tuesday” (123).
This ties quite nicely into the next question, which is a point about alienation, which you talked a bit about before too. Tyler Cowen has this point, for example, that “with perfect information and cleansed preferences, I would prefer red wine to Coca-Cola, but this hypothetical taste has little bearing on my welfare in a world where my preferences are not cleansed” (1993, 262). Again, it would be this worry that laundered preferences would not be suitably connected to people—they would not have motivational force.

The broader concern related to this, also looking at eliciting or using laundered preferences in practice, would be that by invoking people’s laundered preferences instead of their actual preferences, analysts are given quite a few degrees of freedom to decide which preferences should count, which should not count, and we might then have concerns about these analysts imposing their own views on well-being on people instead of deferring to people. Or that some perspectives are removed from the analysis. How would you look at eliciting laundered preferences in practice in terms of who’s in charge of the process and how the process should look like?

On the latter question, obviously the analyst is going to have to sincerely want to elicit the laundered preferences. The analyst needs to have—which is all really a matter of theorizing about well-being—her own understanding of what good information is, what the rationality conditions are, what non-remoteness is, and so forth. She has to sincerely want to elicit that. If she doesn’t, that is an implementation problem—not a problem in theory.

The first part, the alienation problem, is a problem in theory. I believe it is Peter Railton who has a solution.44 Think about it this way. Imagine that either my full-informed self—fully informed, rational, developing non-remote preferences—or my ideal advisor is ranking worlds. Let’s say it is myself fully informed. My fully informed self is ranking possible worlds such that in the possible worlds I may not have full information. So, with full information, I am considering the possibility that in the worlds being ranked, I don’t have full information. Or, for that matter that, in the worlds being ranked, my actual preferences are not the same as my fully informed preferences. Imagine that if I had full information about the process of producing red wine, I would much prefer red wine

44 Railton (1986) proposes an ‘ideal advisor’ account of well-being, which holds that “an individual’s intrinsic good consists in attainment of what he would in idealized circumstances want to want for its own sake [...] were he to assume the place of his actual self” (16).
to cheap beer. But lacking such information, I have no interest in red wine. Now imagine we have two worlds, \( x \) and \( y \), in which I lack information about the process of producing red wine. So in both those worlds I lack the information; in \( x \) I have the wine, in \( y \) I have the beer. My fully informed self would prefer \( y \)—that I have the beer. Because the fully informed self takes account of the actual information of the subject in the worlds. That is, in effect, Railton’s solution and I think it resolves the alienation problem. We are not imagining a ranking of worlds, such that people are fully informed in the worlds, we are simply imaging that the evaluator is fully informed.

The next question takes a leap to interpersonal comparisons of well-being. That is something that is required in the social welfare function approach. One might, however, find your argument in favor of interpersonal comparisons of well-being surprising. We say this because you write in Measuring Social Welfare: “[…] So we shouldn’t accept such an SWF and thus shouldn’t adopt a view of well-being that makes only intrapersonal comparisons” (46); similarly, you write that, “a plausible theory of well-being should allow for intra- and interpersonal comparisons of both well-being levels and differences” (125). This argumentative strategy returns at a number of points throughout the book. Isn’t this reasoning backwards? Surely the correct theory of well-being should be determined independently of any concerns about interpersonal comparability and should then be applied in the framework?

My underlying epistemology here, which a lot of people have when it comes to moral epistemology, is coherentist—it’s reflective equilibrium. Specifically, here, we refine our views about ethical evaluation taking account of our views about the nature of well-being, but we also refine our views about well-being in light of the nature of ethical evaluation. We go back and forth until we have equilibrium. I would say, as a matter of strong intuition, except for a certain corner of economics, both ordinary people (as far as I can tell) and philosophers of well-being accept interpersonal comparisons. There are hypothetical cases: Amy has a high income, good health, lots of friends, an engaged intellectual life; Bob has a low income, is in terrible health, no friends, is bored and miserable. It seems clear that Amy is better off than Bob. Only economists are going to doubt that.

And then, in terms of ethical evaluation, we start with the thought that, if we are welfarists, then the ranking of worlds should depend on
the pattern of well-being. We then find out, when we try to operationalize this for the purposes of a social welfare function, that there are going to be real problems in light of this axiom of ‘invariance’ (which I talk about in chapter two of Measuring Social Welfare), if well-being is not interpersonally comparable. So that is an additional data point in favor of welfare being interpersonally comparable.

Now, you can imagine starting the process of arriving at reflective equilibrium by thinking, ‘I cannot make interpersonal comparisons’. (By the way, why would you think that? You probably think that because you have a preference view of well-being; people have different preferences, how do we make interpersonal comparisons?) You start with a theory of well-being which says I cannot make interpersonal comparisons. You then realize that, in light of this ‘invariance' idea, your social welfare function is going to require an interpersonally comparable well-being measure. At that point, you are out of equilibrium: either you have to give up the notion of a social welfare function; or you have to give up your initial view about well-being, namely, that it is not interpersonally comparable. If that disequilibrium is resolved by changing the view of well-being, I would not view that as being illegitimate—that is simply that nature of reflective-equilibrium reasoning.

The social welfare function approach that you present treats interperson- 
al comparability in a binary way—either we have such comparabil- 
ity, or we don’t—the broader social choice approach, however, does not require us to treat interpersonal comparability in this way. For we can have partial comparability—a whole class of social welfare functions that admits what Sen calls partial comparability.45 What was your moti- 
vation to restrict the class of social welfare functions that you present to the class of social welfare functions that give us some sort of interval or ratio scale comparability, while excluding other social welfare func- 
tions? Is it just a matter of keeping it manageable, or was there some other motivation?

In chapter two of Measuring Social Welfare I have this two-by-two table with admissible well-being comparisons.46 We can talk about either intra- 
or interpersonal comparisons of well-being levels or differences. If a par- 
ticular type of comparison is inadmissible—‘admissible’ or ‘inadmissible’ meaning being eligible or not, recognized or not, by our theory of well-

46 See Adler (2019, 42).
being—then one can never make well-being comparisons of that type. If it is admissible, there still might be pockets of incomparability. For example, assume that I think that well-being levels are comparable both intrapersonally and interpersonally—that is, I think that well-being level comparisons both intra- and interpersonal are admissible. Thus I have a well-being ranking of histories that is not limited to the histories of one person (intrapersonal level comparison) but includes the histories of different persons (interpersonal level comparisons). Now, I am going to say that this well-being ranking is a quasi-ordering—it is a transitive, reflexive ranking of histories. It might be incomplete. It is not pervasively incomplete, it is not massively incomplete, so there is some comparability. Including between the histories of different people. But there could well be pockets of incomparability.

At the end of the day, I think that all four cells in that table are admissible—I think that intra- and interpersonal levels and differences are comparable. I understand that formally—this is what I talk about in my chapter on extended preferences in the Oxford Handbook of Well-Being and Public Policy and also in my appendix to Measuring Social Welfare—as a quasi-ordering of the set of histories (which captures well-being levels) and a quasi-ordering of the set of history pairs (which captures well-being differences), but both those quasi-orderings can be incomplete. 47 There can be pockets of incomparability.

Now, at the level of the social welfare function, what are you going to do? The simple approach, the approach that I use in Measuring Social Welfare, the approach generally used in the literature, is to assume that there is a single well-being measure. For my purposes, a single well-being measure that captures levels and differences. That presupposes completeness—if there is incompleteness in the comparisons of either levels or differences, the levels ranking and differences ranking cannot be captured by a single well-being measure. The simple approach is for tractability. There are ways to generalize the social welfare function beyond this operational assumption of completeness, which I discuss in my 2012 book. For example, in that book, what I do is to think of the well-being comparisons being represented by a set of well-being measures. Well-being comparisons of levels and differences correspond to a super-valuation over that set and the social welfare function is applied to all the well-

being measures in that set. This is going to be a way to allow for incompleteness within the social welfare function framework. But, again, it is simpler, more tractable, to just assume a single well-being measure.

*The general point, if we are understanding this correctly, is that you are sympathetic to having a more fine-grained distinction between incomparability at one end—intrapersonal incomparability—and full ratio scale of comparability at the other.*

Correct. Well, there are two issues. One issue is, is well-being measurable on an ordinal scale, an interval scale, or a ratio scale? The question there is, can we make comparisons of levels, comparisons of differences, and/or comparisons of ratios? I actually think that we can make comparisons of all three. Then, in making level, difference, and/or ratio comparisons, are those complete or not? There I would say ‘no’. Plausibly, there are going to be pockets of incomparability in making level comparisons, difference comparisons, or ratio comparisons. We are going to be somewhere in the middle between full completeness and global incomparability. That is going to be the best theory of well-being and, again, just in terms of making the social welfare function approach implementable, we are going to assume, or impose, completeness by having a single well-being measure. But that is a simplification.

*A related, but also distinct question, on the metaphysics of value: some of these approaches require strong scaling assumptions, which you also spoke about earlier. Minimally you normalize a lower bound at zero. How do you interpret that? Is that the worst possible life that you can lead, or is that just neutral between positive features and negative features?*

It depends. Utilitarianism in the fixed population context only needs a well-being measure which is unique up to a positive affine transformation. It does not really matter what the zero is. That becomes different in the variable population context. In a variable population context, it depends on what the variable population view is: there is total utilitarianism, there is also critical level utilitarianism. If you are a total utilitarian, who says adding someone to the population is good as long as that person is better off than in the neutral life, it is natural to assign zero to the neutral life. If you are a critical level utilitarian, who says—this is Broome’s view—there is some critical level that is better than neutrality, such that adding someone to the population is only good if the person is above the critical
level—this is designed to avoid the repugnant conclusion—then it might be natural to assign zero to the critical level. You can do either one as long as you are aware of what you are doing.

For prioritarianism, it is more complicated. No prioritarian social welfare function, whatever the concave transformation function might be, is going to be invariant to a positive affine transformation of the well-being measure. With the Atkinson prioritarian social welfare function, even in the fixed population context, the interval scale of well-being is not enough. We have to fix a zero point. That zero point is the Atkinson zero, in the sense that it has got to be no better than, and probably worse than, any of the lives under consideration. The problem with Atkinson is that it does not work with negative well-being numbers. We are going to have to pick zero so that none of the lives under consideration are assigned a negative well-being number.

IV. THE ETHICAL DELIBERATOR AND THE ROLE OF THE PHILOSOPHER

**Changing track a little bit.** There are a number of value judgments central to the social welfare function approach, as you discuss in Measuring Social Welfare: “Choosing a correct theory of well-being, constructing a measure w(.), adopting a rule E, specifying an uncertainty module” (79). How are these questions to be answered? Your own answer to this consists in presenting reflective equilibrium as a moral epistemology and you write: “It is for each ethical deliberator to locate her own point of reflective equilibrium—to determine which SWF she favors, on balance and after reflection” (115), that is, which “SWF most fully embodies her ethical commitments” (96). Can you elaborate on how this is supposed to work in practice?

In practice, at the level of each of us as individual thinkers trying to decide what we think morally, the situation is just what I say. You have to first of all deliberate about whether you are inclined towards consequentialism or some kind of non-consequentialist view, say a deontological view; within consequentialism, towards welfarism as opposed to non-welfarist consequentialism; and then within welfarism, you need to think about what welfarist world-ordering you favor—that is going to be a matter of specific thought experiments, we can have thought experiments which are

---

48 See Broome (2004); see also Blackorby, Bossert, and Donaldson (2005).
trouble for utilitarianism—but one of the nice things about the social welfare function literature is that it also shows the role of axioms in thinking about different world orderings. And again, E is a world ordering, more precisely it is a way to express the ranking of worlds in light of well-being. In choosing between the utilitarian E rule, the prioritarian E rules, egalitarian, leximin, and so forth, thinking about how those relate to different axioms is very helpful. And, indeed, what I talk about in chapters three and four in *Measuring Social Welfare*—here I am just synthesizing the existing literature—is that if we use the five axioms of Pareto, anonymity, Pigou-Dalton, separability, and continuity, we can really narrow down the space of different social welfare functions and crystalize our thinking. Most people in this space find Pareto and anonymity pretty compelling, I also find Pigou-Dalton quite compelling. Pigou-Dalton is, of course, the axiom that separates between utilitarianism and prioritarianism. So, part of the way that I come to my own, prioritarian, reflective equilibrium is by thinking about and endorsing Pigou-Dalton.

*The follow-up question to that is how this idea of the ethical deliberator will play out in practice, who this ethical deliberator is supposed to be when we apply the social welfare function framework?*

That is more complicated. The complication is that, if ‘in practice’ means in government, then we have to think about the relation between law and morality. Because government is constituted by different officials with different official roles. And each official is given legal powers and restrictions, and what an official does is constrained by those. An official cannot simply do what she thinks is the ethically best thing—she has to act within a role. For example, in the United States, we have civil servants in the federal government who are not elected. And I think a lot of people would think it would be legally problematic—not ethically problematic, legally problematic—for an unelected civil servant to decide, 'I am just going to operationalize whichever social welfare function I like as an ethical matter'. Rather, that seems to be the kind of judgement which legally should be made only by a legislator or, in the U.S., by the elected President. As an individual, or as a citizen, I am not legally constrained; I can act on my own social welfare function in voting or donating my money. The additional complication comes in to the extent that we have officials with a legal role and operating under legal constraint. There again we have to think about the connection between law and morality, which is a complicated topic. But here I would say at least that, in a democratic legal
system, there are going to be some decision makers with sufficient democratic authorization to be able to choose between different social welfare functions. At a minimum, that would be legislators; a legislator can pick a social welfare function to help him decide how to vote on proposed legislation. And there might be other officials—for example, in the U.S., presidents—operating within their zone of legal discretion.

The upshot would basically be that, whether we should have a prioritarian or utilitarian social welfare function is ultimately something that should be democratically informed as you see things?

Well, yes and no. I am a citizen. Let us imagine that everyone else in the country is utilitarian and I am prioritarian. I should not necessarily change my view—I might think everybody else is misguided. I might change my view: the fact everyone else is utilitarian might lead me to reflect, but if after reflection I still believe in prioritarianism, I have no democratic obligation to change my view simply because everyone else is utilitarian. Again, we are welfarists. I hate to say this, but it is the case: welfarists believe that democracy has no bedrock moral role. That itself is complicated because you might say that democratic values are somehow related to well-being. But at least to a first approximation, democratic legal structures have no bedrock ethical role for welfarists. It is rather that democratic legal structures are a good way in practice to advance well-being—look how dictatorships are doing as compared to democracies. So, the point is that we have good welfarist reason to promote democratic legal structures, and part of democratic legal structures is that people’s legal roles are limited and that officials without democratic authorization cannot legally make certain decisions that officials with democratic authorization can. But again, it seems to me that—and this is a very complicated topic—all of these democratic constraints are a matter of law, not morality, and that the welfarist—whether she is a utilitarian or prioritarian—sees democratic legal structures simply as a useful instrument to promoting well-being.

Relatedly, how do you see your own role and the role of philosophers in the social welfare function approach? Is it to make the strongest possible case for particular views on these substantive (ethical) issues (that are involved in choosing these different functions, a well-being measure, and uncertainty modules)? Or is it to provide a ‘taxonomy’ of alternative specifications without really taking any side? We ask because
the book can be read as laying out the framework (assumptions and choices) and providing reasons for specific social welfare functions without really taking a side. Nevertheless, your own sympathy for the prioritarian class of social welfare functions does come through as well.

The growth of knowledge over the last century, let us say, or 150 years—whatever it is—about both economics and, within philosophy, about consequentialism, has been unbelievable. We have a much, much better formal and substantive understanding of these subjects—I mean, think about where philosophy was, say, before Sidgwick started writing, before the 19-century utilitarians started writing; let alone the development of economics with Samuelson, Arrow, Sen and all that, I mean the explosion of knowledge—genuine knowledge—has been fantastic. So, I think philosophers can do their bit as part of that and that can be different things. It really depends on what you are good at and what you want to do. One thing is to present the pros and cons of different approaches and develop our understanding of possible axioms, the implications of those axioms, ways to characterize different views, and so forth. It is also valuable to advocate for a particular view. For example, think about Larry Temkin; his book Inequality is a great book. He has devoted his academic efforts to arguing for the value of equality. That has made a huge contribution. Contrast John Broome. Broome is a utilitarian but it is not as if Broome has argued in a full-throated way for utilitarianism in the same way that Temkin has argued for equality. Rather, Broome has been exploring the axiomatic and substantive arguments for and against utilitarianism. I think Broome has made an incredible contribution. So, it is really what you can do. My advice would be to philosophers, work on problems that engage you and make the contribution where you can make it.

They also feed into each other, right. We also posed the question in a rather stark way. Clarifying axiomatic structure can through reflective equilibrium clarify for you what your position is, and the other way as well, Temkin is an instructive example here—the advocacy can also clarify what is at stake.

Consider Parfit. In the fourth part of Reasons and Persons, he identified some very important puzzles for population ethics—the non-identity problem, the repugnant conclusion—and a key contribution was to say, ‘I cannot identify a theory that avoids all these unpleasant consequences; I

49 See Sidgwick’s The Methods of Ethics (1874).
50 See Temkin (1993).
would like to come up with a theory X that does that, but I haven’t yet done so’. That intervention has been incredibly fruitful—he launched an entire generation of people looking for theory X.

On the other hand, if you think about part three of *Reasons and Persons*, where Parfit adopts a particular view of personal identity—kind of a psychological continuity view—there he is less skeptical, the idea seems to be, ‘this is what personal identity consists in’. That has also been very valuable in the literature on personal identity—people react to him. That is less skeptical, or less open, but it has also been very useful.

*This ties in very well with the last question of this section. It will be clear to everyone who reads* Measuring Social Welfare, and your other work as well, *that John Harsanyi and Derek Parfit have had a significant influence on your thinking. Can you tell us about how you came across their work and what the nature of their influence has been on your thinking?*

Even though I did do a fair bit of philosophy in college, I don’t think I read either of them during college. I did grad work in history—didn’t read them then. I was at Oxford, but I never met Parfit when I was there as a grad student in history. It’s when I started law teaching, teaching seminars on the ethical foundations of cost-benefit analysis, that I started to read both Harsanyi and Parfit.

Parfit’s *Reasons and Persons* is widely seen as perhaps the greatest work of moral philosophy over the last century—certainly the greatest work of moral philosophy in the utilitarian or welfarist tradition over the last century—so it has to be read and grappled with. As for Harsanyi: with reference to cost-benefit analysis, I was very interested in this problem of interpersonal comparisons. As mentioned, the view that I was developing with Eric Posner was that we should think about cost-benefit analysis as a rough proxy for overall well-being. I was also inclined towards a preference-based view of well-being. So then the puzzle was, ‘how do we adopt a preference-based view of well-being and allow for interpersonal comparisons?’, and of course Harsanyi proposes a solution to that. That’s really it. I came to Harsanyi’s work on extended preferences as a result of my interest in both preference-based views of well-being and interpersonal comparisons.
V. LOOKING AHEAD

What’s next for you in terms of your research and the social welfare function approach? What future developments are you looking for in terms of the social welfare function approach in both theory and application? What research questions are you looking to tackle in the future?

On the more practical side, we just finished this Prioritarianism in Practice book, which involved a whole team of scholars—with different chapter authors looking at the application of prioritarianism as contrasted with utilitarianism and other standard approaches (cost-benefit analysis, cost-effectiveness analysis), in different policy domains. The policy domains included climate policy, optimal taxation, risk regulation, health policy, COVID-19, education policy.

I am spearheading with Ole Norheim a follow-on effort, which we are calling PiP Health, continuing the work on the application of prioritarianism in health policy. QALYs and QALY-based cost-effectiveness have had a big role in health policy assessment in various countries. But for these frameworks, the measure of individual attainment is health. The QALY approach uses a measure of health, which combines longevity and health states, and then the impact on income is considered separately. That is cost-effectiveness analysis. What PiP Health is trying to do is thinking about health policies using a measure of lifetime well-being that integrates longevity, health, and income. And then using that as the input into a prioritarian social welfare function. The chapter on health policy in the Prioritarianism in Practice book does this already, but we are hoping for follow-on work and other related work on prioritarianism in health.51 That’s the applied project.

The other thing right now: it may be that the United States is going to move to, finally, incorporating distributional weights into cost-benefit analysis—distributional weights are a way to approximate a utilitarian or even a prioritarian social welfare function.52 So, I may be involved in that effort, or writing about that effort.53

More theoretically, I think there are a whole bunch of outstanding questions still for prioritarianism. For example, prioritarianism and non-
human animals. How we should take account of the well-being of non-human animals if we are prioritarians is a really important topic. There has been a broader interest in social welfare functions and animals, so that is something that I hope to write about.

There are other pieces of the puzzle that I think, philosophically, still need to be worked on. For example, the whole issue of separability. Separability is this technical axiom that distinguishes between, as I see it, prioritarianism and a kind of moderate egalitarianism. A moderate egalitarian social welfare function respects Pareto, anonymity, and Pigou-Dalton, but fails separability, while prioritarianism respects Pareto, anonymity, and Pigou-Dalton, and satisfies separability. I think more philosophical attention needs to be given to separability, and there are lots of complicated issues about separability not just under certainty in the fixed population case, but with variable population and under uncertainty. I’m hoping to work on these issues.

A third thing I’m working on now, in terms of philosophical fundamentals, is luck prioritarianism: the question of how to incorporate considerations of desert and responsibility into prioritarianism. The thought behind luck egalitarianism and this whole economic literature on equality of opportunity is that, at the end of the day welfarism is insufficient—because we should care about the extent to which people are responsible for being badly off as opposed to being badly off through no fault of their own. How to make sense of that thought under the rubric of prioritarianism is an open question. I took a stab at it in an article in *Utilitas*, in 2018—“Prioritarianism: Room for Desert?”—but much more works needs to be done.54 A final piece of the puzzle is the debate between luck egalitarians and relational egalitarians. Relational egalitarians—I am thinking here of the work of Elizabeth Anderson, for example, or more recently of Juliana Bidadanure—are non-consequentialists who care a lot about social relations, or social status, as opposed to the distribution of income or well-being.55 I think one question for, let’s say, prioritarian welfarists is how we take account of relational egalitarian insights. The key relational egalitarian insight, as I see it, is that, to put it in welfarist terms, social status matters a lot for well-being. That the relations of social hierarchy, oppression, subordination—which of course we continue to see all around us—have a huge impact on individuals’ well-being. The question is, what exactly does that mean as a prioritarian—how do we take account of these

54 See Adler (2018a).
55 See, for example, Anderson (1999) and Bidadanure (2016).
problematic types of social relations within prioritarianism? Do we somehow funnel them into our well-being measure? How is that going to work? It has been avoided in practice, because typically when prioritarianism is implemented the inputs to well-being are things like income, longevity, health, and so forth, and not the quality of someone’s social relations or social status. There are a range of issues, both theoretical and practical, that need to be engaged there.

And then the other thing I am doing, I am finishing up this book on Risk, Death, and Well-Being: The Ethical Foundations of Fatality Risk Regulation, looking at fatality risk regulation from a prioritarian as well as a utilitarian perspective. All of this will keep me busy over the next couple of years.

*Your academic work is, as we have explored, highly interdisciplinary in nature. How do you look at the future of interdisciplinary work in law, economics, and philosophy? What developments are you seeing and what developments are you hoping for?*

I think there is a lot of good work occurring at the intersection of economics and philosophy, certainly. In the normal way, generated either by problems that are coming up in the literature or externally. I can think of a number of examples. One is all of this work on variable population. The issue of variable population, or population ethics, has always been a fruitful source of interchange between economists—for example, Blackorby, Bossert, and Donaldson—and philosophers—obviously going back to Parfit or Gustaf Arrhenius—and that continues to happen.56 We have a new generation working on these problems, people like Dean Spears, Mark Budolfson, and others from the economics side and philosophers as well.57 So, that is one exciting area. There are a lot of people working both on different possibilities and on advancing challenging new ideas; for example, an increased willingness perhaps to accept the repugnant conclusion. In any event, population ethics is one area where I see a lot of new work at the intersection of economics and philosophy.

The other thing right now, growing out of effective altruism, is the whole interest in long-termism, fanaticism. The idea that we should devote a lot of efforts to increasing the chance that humanity continues for a very long time, and should be willing to impose present costs for the

---

57 See, for example, Spears and Budolfson (2021).
sake of only a small chance of increasing long-term survival, has generated this literature on fanaticism—coming out of, for example, the Oxford Global Priorities Institute and other such think-tanks, and including publications in *Ethics*. This is also proving to be a very fruitful area for work at the intersection of economics and philosophy.

A third example is the work now among contractualists about risk regulation. How contractualists growing out of Scanlon’s work on contractualism should think about permissible risk imposition. It is philosophical but it is also at times fairly formal, and I view this as another exciting area of research that integrates formal methods with philosophy.

Topics for collective attention emerge organically in the literature, which is the way scholarship works. And then we have external stimuli; for example, for a long time now thinking about climate policy has been an important external stimulus to economic and philosophical work.

*The EJPE is an interdisciplinary journal, and our readers are scholars who either do work at the intersection of philosophy and economics, or are at least open to such an interdisciplinary approach. What advice would you give to graduate students aiming to pursue an interdisciplinary academic career, such as your own?*

I feel that I have been extremely lucky. The U.S. legal academy is very open to interdisciplinary approaches; I was fortunate to get a job at a U.S. law school and thus to have the freedom to pursue my research interests, educate myself in different fields—again I came in without a PhD in either economics or philosophy and was able to pursue those interests. As a matter of practical advice: the first thing is to get a job. If you want to be able to pursue interdisciplinary scholarship over a lifetime, the first thing is to figure out, depending upon your country, what sorts of long-term positions are available. In the U.S. that is heterogeneous—there are positions at law schools, there are positions at public policy schools, there are positions at philosophy departments, there are positions at econ departments—so think about that, and think about your own skill set. Earning a PhD or another advanced degree (in my case, a law degree) and landing a job is a twofold thing: on the one hand it is about pursuing your interests, on the other it is an exercise in finding a job and then, later, getting tenure. So, again, the most pragmatic advice is to keep in mind employment prospects as you pursue your studies.

---

Another point is that work at the margins of economics, law, and philosophy can proceed in lots of ways. It can be quite formal; it can be quite philosophical and conceptual; it can be focused on legal implementation, which is very complicated. So, figure out which of those you enjoy most and which of those you are most able to contribute to. If you don’t have a lot of interest in, or aptitude for, math then don’t try to do really formal work—do philosophical work and try to collaborate with formal people. And vice versa: if your value-added is the math side, that’s great, try to do the philosophy too, but also be prepared to collaborate.

In short, it’s a matter of figuring out where the jobs are and what it takes to get a job, figuring out what interests you, and figuring out what your skillset is.

Now, if you are fortunate enough, as I have been, to have a tenured job where you can pursue a bunch of research questions, at that point, the issue is not landing a job, but it is still figuring out where you can have an impact, which also relates to what interests you and what your capabilities are.

As a final question, if you had to name three philosophical works that any graduate student in law, economics, and philosophy should read, which would those be?
I’d certainly say Parfit’s Reasons and Persons.\(^{59}\) As I already said, the greatest work in moral philosophy—certainly welfarist moral philosophy—in at least the last century. And Broome’s Weighing Goods.\(^{60}\) Broome has been incredibly important in this field. It is an attempt to synthesize economics and philosophy, by someone who was first a chaired professor in economics then a chaired professor in philosophy. So, Reasons and Persons and Weighing Goods.

Now picking the third is a little tricky. I think there are a number of possibilities. Sen’s Collective Choice and Social Welfare is incredibly important, it is more formal than philosophical.\(^{61}\) I think that everyone needs to have read Rawls’ A Theory of Justice.\(^{62}\) Two other books that I think are really important. The best single book on the philosophy of well-being is Sumner’s Welfare, Happiness, and Ethics—it’s a great book.\(^{63}\) Harsanyi did incredibly important stuff, but he didn’t write a book-length work that I’d

---

\(^{60}\) See Broome (1991).
\(^{62}\) See Rawls (1971).
\(^{63}\) See Sumner (1996).
recommend. I would consider Temkin’s book *Inequality*, a really important defense of the value of equality but, more generally, just incredibly rich philosophically.\(^{64}\) I realize this is two plus a bunch as opposed to three…

*Professor Adler, thank you for sharing your time and ideas with us.*

Thanks a lot, I appreciate it.

**REFERENCES**


---

\(^{64}\) See Temkin (1993).


JOSEP RECASENS  
*Universitat Pompeu Fabra*

The arc of the moral universe bends towards justice, but there are certain situations where it seems more urgent to bend it. One of them is when someone has been wronged and a reparation is needed. Another one is when a disaster occurs. In *Reconsidering Reparations* (2022), Olufèmi O. Táíwò tries to establish the moral link between the past wrongs of colonialism and the already present disasters of climate change.

The main thesis of the book suggests that reparations for colonialism must be able to assure everybody's capabilities to have relationships of equality with others. Táíwò names this project the ‘constructivist’ view of reparations. Since climate change is distributing advantages or disadvantages in an unfair way, the constructivist project requires that climate justice be successful. The book has 6 chapters. Chapter 1 is an introduction. Chapter 2 explains how the current world system is a direct consequence of colonialism. Chapter 3 presents Táíwò’s constructivist view, and Chapter 4 confronts it with two alternative reparations approaches. Chapter 5 shows the connection between reparations and climate justice. Finally, in Chapter 6, there is a reflection on the responsibility of intergenerational justice.

The book is divided into two parts upon which this review will briefly comment: reparations for colonialism and climate justice. The first part begins with the exposition of the historical situation in Chapter 2: the current distribution of advantages and disadvantages within the present-day global economic and political framework (the Global Racial Empire). This can be directly traced back to the historical injustices stemming from the Atlantic colonialism that emerged following 1492 (24). The differences between countries' advantages and disadvantages have this historical origin, but continue to determine them because advantages are cumulative (25).

The existence of historical injustice seems to imply that this wrongdoing needs to be corrected by the wrongdoer; that is, corrective justice is required. Previous reparations accounts such as the “harm approach”
(124) or the “relationship repair view” (133) assume this perspective. However, Táiwó argues that these backward-looking accounts suffer from different versions of the non-identity problem (127): the people that caused the injustice are dead and no living people can be held accountable for those acts. Furthermore, corrective justice can be applied and still leave an unfair distribution of advantages.

The problems of corrective justice allow Táiwó to present his ‘constructivist view’ of reparations as a distributive justice approach. According to this account, what is unjust is the unfair distribution of advantages and disadvantages that started with colonialism and continues in present structures (75). Therefore, what needs to be done is to modify these structures by “remaking the world” (72). This view is defined by forward-looking targets (it resembles Young 2006). It does not seek to punish wrong-doers but intends to correct the unequal distribution. However, it is historically informed when assigning the costs of this project (73).

More specifically, the **distribuenda** of this approach are the capabilities needed to function as an equal and free individual (inspired by Nussbaum 2011). Táiwó considers advantages or disadvantages as the elements that expand or shrink someone’s capabilities (88). This not only includes material resources, but also social norms such as racism.

This historical fact has a crucial importance for the debates about the limitations of Global Justice. The first refers to the duties of distributive justice the citizens of one state have to their fellow citizens, and the question about whether they are different to the ones they have towards non-members. Some argue that when two countries have different wealth levels, if the richest state has no interaction with or responsibility for the other state’s situation, the duties of justice to members or non-members of their state are not the same (see Miller 1995). The second debate includes the problem of brute-option luck (Dworkin 2002). If two states begin with the same amount of money but one gets rich by investing it in technology and the other poor because it wastes it, does the rich country have duties of distributive justice towards the poor one?

By situating his reparations approach in this historical context Táiwó avoids both limitations to his project of Global Justice. The constructivist view must be Global because the whole world is shaped by the reality of colonialism. Nonetheless, he only focusses on Atlantic history (84) and leaves open the issue about the geographical limitations of this approach. Also, the cumulative aspect of advantages makes especially poor states not responsible for their situations (86).
Let us comment on the second aspect. Táiwò argues that the relation between colonialism and climate change is “contingent” (158). The explanation of this idea is useful to understand the role climate justice has in the constructivist view. It is true that the countries that will suffer the greater effects of climate change tend to be the ones that were colonized. This is, in part, an unfortunate coincidence due to natural processes. However, the impact caused by climate change also depends on the level of adaptation, which is higher when there are material and intellectual resources to prepare for natural disasters (as seen in the example of Hurricane Katrina’s effects in New Orleans (153)). Once we understand that reparations must deal with the distribution of advantages and disadvantages, it is easy to comprehend that climate change adds new advantages and disadvantages (“environmental risks and vulnerability” (167)) that were not foreseen during the history of colonialism. Therefore, since the inherited structures of colonialism are the ones that distribute the levels of prevention throughout the world, it is necessary to include climate justice in the constructivist view.

This book tackled two main difficult tasks. The first one was to define a project of reparations for colonialism that was sound and philosophically coherent. This is well accomplished because it embraces a forward-looking rather than a corrective approach. However, there are two issues that have not been addressed in the book that would clarify the author’s position. The first one is the importance of colonialism for the justification of a distributive justice project that wants to distribute capabilities equally. Is there a duty of justice to do that even if inequality was not caused by colonialism? Second, when Táiwò says that the constructivist view can be compatible with other corrective justice approaches (124), he does not explain in which ways. This absence makes it difficult to understand the real connection between Táiwò’s account and the correction of the wrongdoings of colonialism. Can corrective and distributive accounts of reparations really be compatible (see Blomfield 2021)? Should the advocates of reparations embrace both accounts if they are compatible? Would this increase the specific demands when both kinds of justice apply?

The second task was to convincingly link the reparations for colonialism to the climate justice project. This is the most original part since nobody had argued for this connection. However, the book does not fully capture the relationship between climate change and colonialism. One of the main omissions is the explanation of the economic processes required
to increase capital accumulation and the environmental damage they cause. Pollution is not merely a side effect of colonialism, but a necessary element to maintain the inherited structures of advantages and disadvantages, as Malm (2016) shows.

In conclusion, Reconsidering Reparations introduces new intuitions to the usually philosophically stagnant debates of reparations and climate justice. This book will be of interest to scholars and general readers interested in the philosophical justification of progressive politics. Specifically, it can be of great help for those who know that we need to consider the injustices of the past to build a just and livable future.

REFERENCES

Josep Recasens is a PhD student at Universitat Pompeu Fabra. His research interests include compensation for Loss and Damage, insurances and climate justice. He holds a BA in Philosophy from the University of Barcelona and a MA in Political Philosophy from the Universitat Pompeu Fabra.
Contact e-mail: <josep.recasens02@estudiant.upf.edu>

BOB R.W. VAN VELTHOVEN  
*Leiden University*

With the provocative title *The Greeks and the Rational: The Discovery of Practical Reason*, Ober chooses two fundamental works as his reference point: E.R. Dodds’ *The Greeks and the Irrational* (1951) and Bruno Snell’s *Die Entdeckung des Geistes* (1946). Whereas Dodds famously argued against Greece presenting a ‘triumph of rationalism’ by pointing out pre-modern attitudes and modes of thinking, Bruno Snell asserted that the Ancient Greeks in Homeric times lacked a conception of the self and later gradually discovered introspection and a theory of mind. Following the line of his earlier books, Ober argues the exact opposite: the Ancient Greeks did not only behave rationally but also had a discourse about various aspects of rationality. The general claim of Ober’s book is that a so-called ‘Folk Theory’ of instrumental rationality emerged in the Greek world in the first half of the fifth century BCE. Although this Folk Theory does not resemble the form of modern rational choice theory, it made similar assumptions, such as that people employ their cognitive capacities to form ordered preferences and consistently act to obtain the best possible outcome. Ober (6) illustrates this by quoting Xenophon, *Memorabilia* 3.9.4: "that all persons deliberately choose, out of what is available to them, what they think is most advantageous to themselves, and they do this".

This theory of instrumental rationality was first propagated by the Sophists and, through their education, their ideas permeated Greek public discourse, leading to numerous reflections in the late-fifth and early-fourth-century Greek philosophical and historical texts. When this popular theory was brought into practice in ancient Athens, it led to a process of institution formation (chapter 4) that enabled the Athenian democracy (chapter 6) to become a rational and, therefore, dominating force in contemporary international politics (chapter 5). Moreover, when practiced by individuals, it led to remarkably high (for premodern standards) economic growth (Chapter 7; see also Ober 2015).
While the successes of their conceptualization of the Greek individual as "unerring craftsmen of self-interest" (Plato, Republic 340e–341a; Greeks and the Rational, 46–47) are visible in retrospect, instrumental rationality was, as Ober shows, always met with suspicion in the Classical Era. As Ober shows in the first three chapters, authors such as Plato and Herodotus feared that pursuing self-interest, although considered a foundational force in the birth of political communities, would become detrimental to society. In Chapter 1, Ober describes how, according to a ‘Thrasymachean’ variant of the folk theory, the strong by nature (phusis), who are subjugated by the weak through convention (nomos), would become criminals and tyrants if their restraints were lifted. In chapter 2, Ober shows how even according to the ‘conventionalist’ variant of folk theory, in which individuals are willing to form communities for protection, the fear of Thrasymachean defectors is a challenge that needs to be overcome. In Chapter 3, Ober shows how ‘constitutional moments’ (such as successful coups and revolts) that demand a proper democracy—the constitution that will best serve the people’s interests—can be hijacked by aspiring autocrats. These concerns led Plato to develop what Ober calls a theory of “ethical rationality” (135), which combines the strength of means-to-end reasoning with the postulation that the ends of reason must also be objectively reasonable (chapter 6).

As Ober concludes in chapter 8, by reinterpreting Aeschylus’ tragedy Eumenides, the fifth-century BCE Greeks first formulated a theory of practical reason. For the fourth century BCE Socratic philosophers Plato, Xenophon, and Aristotle, the rediscovery and ethical refinement of this theory was a central effort, which ultimately led to Aristotle’s claim in the Eudemian Ethics and Nicomachean Ethics that for a man and a state to achieve happiness (eudaimonia), instrumental means-to-end reasoning must be combined with a virtuous choice of ends. In the line that Ober draws through Western philosophy, this idea of eudaimonistic ethical egoism was dominant until the 18th century, when it was challenged by Kant’s deontology (egoism is not ethical) and, to a lesser extent, utilitarianism (prudent egoism does not need to be unethical). However, according to Ober, the Socratic ideal was revived by John Rawls, who emphasizes that the ideal moral agent should both be (means-to-end) rational and reasonable in his choice of ends. Ober concludes, in opposition to several of Rawls’ many critics, that this Socratic ideal can inspire a new approach combining positive and normative approaches to political theory and thereby overcome the limitations of rational choice theory and the hazard
that normative considerations are crowded out by instrumental reasoning (following Amadae 2016).

In his book, Ober takes his readers on a tour de force through history and philosophy, departing from Ancient Greek ideas about the role that self-interest plays at the very origins of society and showing how the Ancient Greek analysis of the problem of “vicious egoism” can be still relevant today (392). The journey is layered, rich, and generally well-anchored in multiple debates, as the carefully composed introduction clarifies. While Ober's argument makes existing lines of research on rationality in ancient Greece more robust, it adds much by taking the Greeks’ discourse (rather than their behavior) as the central unit of analysis. Throughout his argument, Ober shows that he is an attentive reader, a sharp analyst, and a rigorous reasoner who makes this book a true conversation partner for social scientists.

While reading and discussing the book with fellow classics scholars interested in Ancient Greek economic performance and theory, I noticed that Ober's argument is complex, and the book’s premise is too quickly criticized and not considered carefully. It, therefore, seems a good starting point to defend the book from some clear misconceptions. Ober's claim is not that the Greeks were rational *per se* (no more or less rational than we are, with well-known cognitive biases and limitations); that they were the first to be rational (they were not); that they were the first to discover rationality (they were not); that they only became rational after discovering it (Ober compares it to "Molière’s disingenuous M. Jourdain, who, under the tutelage of a ‘master of philosophy’, discovered to his amazement that he had been speaking prose for his entire life" (1)); or that they had the first discourse about rationality (they did not: Ober relativizes the uniqueness of Greek discourse and suggests that, e.g., Indian and Chinese discourses could as well be examined (3–4)).

Another potential point of criticism, which Ober addresses himself, is his use of game theory to explain how some narratives or situations are rational. This approach is relatively new in the humanities (38–40). Among others, Ober (2015) has pioneered its application to Classical Studies, which has the potential to produce more promising work in the future (cf., Węcowski 2022). While Ober himself concedes that this method, including his focus on the language of choice and reasoning, is reductive because it leaves many contexts out of consideration (15), I think Ober's careful and insightful application of game theory shows this
approach can be rather revealing of elements we would otherwise have missed.

For example, in chapter 3, Ober analyses the voting passage in the so-called ‘Constitutional Debate’ (Herodotus, Histories 3.80–82), in which seven victorious putschists decide between three constitutions (democracy, oligarchy, monarchy) for the Persian empire. Ober demonstrates that if such a decision with this setup had been made through multiple rounds of voting, Darius, the later Persian King who had pleaded for monarchy during the debate, is likely to have profited from a weakness in the voting procedure by changing his vote in between the rounds, thus manipulating the result to his advantage. Herodotus, as Ober innovatively argues, may have used this story to show that the decision-making mechanisms of democracy can be fickle and will not always produce a rational result that fairly aggregates the voters’ preferences.

Another misconception that one could have about this book is that it promotes rational choice theory, which has become the object of criticism for promoting an inaccurate and harmful image of human nature (Bowles 2016; Urbina and Ruiz-Valverde 2018). Although Ober only mentions some criticisms of rational choice theory at the end of his book, it becomes clear that he sees it as a model and means of theorizing that is at times a fallible predictor of an opponent’s action. In chapter 5, Ober describes how Thucydides theorizes that Pericles’ foresight and his ability to take risks upon accurate calculation make the Athenian empire thrive at the offset of the Peloponnesian war. Whereas Athens may continue to function as a rational agent after Pericles’ demise if it has prudent leaders (and good institutions), city-states such as the much weaker Melos may act irrationally and revolt against the Athenian might because of the prospect of being free (Kahneman and Tversky 1979), leading to their annihilation.

Chapter 7, on which I now would like to focus, plays an essential role in Ober’s effort to show what rational choice theory leaves out if it is divorced from practical reason. Whereas a considerable part of scholarship in this field collects evidence for rational, maximizing economic behavior in order to refute Finley’s (1973) idea that the Greek status-focused mentality prevented them from economic thinking (cf., Leese 2021), Ober departs from this line of reasoning by showing that ancient ‘entrepreneurs’, such as Cephalus in the Republic, do not act as ideal types of economic rationality (297) (contrary to his brief treatment of economic rationality in Aristotle and Xenophon later on). Using game theory, Ober
demonstrates how, in the *Iliad*, the greedy leader Agamemnon is a “clueless” game player and how he worsens the epic’s central conflict by switching between his unranked preferences (314ff.). Arguing against Snellian explanations, Ober shows that Agamemnon serves as a foil, undermining the ideal of the leader as a rational and reasonable decision-maker (since Achilles characterizes Agamemnon as someone "that does not at all know to look both backward and forwards in time" (*Iliad*, 1.343–344; (310)). Ober connects this point to Socrates’ conversation with Cephalus and Polemarchus about the importance of being equitable (*epieikeia*) in the *Republic*, a virtue that, in economic reality, enabled instrumentally rational people from different social classes (including the enslaved) to engage in commercial exchange with each other.

This chapter shows the strengths and limitations of Ober’s approach. First, Ober innovatively refutes Finley’s line of argumentation by demonstrating that the Greeks were capable of instrumental rationality and possessed economic mentality. However, he then shows, based on the normative philosophical discourse, on which also Finley had based his argument, that successful commercial practice (in theory) requires both instrumental and value-driven reasoning. Second, Ober’s focus on discourse rather than performance balances the widely used New Institutional Economics (NIE) approach. However, his (insightful yet indeed reductive) reconstruction of a single discourse on rationality leaves little room for appreciating discursive pluralism (380), whereas Xenophon’s oeuvre and works belonging to the genre of *oikonomia* literature could have formed promising opportunities to do so (Helmer 2021; Hinsch 2021). Furthermore, Ober’s re-evaluation of practical reason could have been supported by delving into ‘thicker’ notions of rationality. Fruitful concepts to work with could have been *aspiration* (Cephalus’ aspiration to bequeath more than he had inherited himself both informs and limits his instrumental reasoning—see Wincewicz 2018) or *ecological rationality* (Agamemnon’s failure at strategic rationality shows the value of correct intuitive decision making—see Todd and Gigerenzer 2012). Most importantly, however, while Ober provides evidence for the discourse on rationality and the astonishing economic performance of ancient Greece, the precise connection between the two often remains tentative. Hopefully, this relationship will be developed in future research.

Apart from hopefully becoming a landmark publication and a source of inspiration for many classics scholars, ancient historians, philosophers, and other humanities scholars, this book promises to be an
intriguing read for any political or social scientist working on game theory and rationality in theory and performance. For such scholars, this perhaps unexpected engagement with the Classics can be surprisingly stimulating and insightful. As Ober states: "Reading classical texts with and against formal theory alerts us both to the dangers of rationality divorced from ethics and to the unworldliness of politics without strategic rationality" (392).

REFERENCES

Bob van Velthoven is a Ph.D. student in Classics at Leiden University. His dissertation focuses on economic theory and the conceptualization of economic rationality in Socratic philosophers. He is part of the NWO VIDI Project ‘From Homo Economicus to Political Animal’.
Contact e-mail: <b.r.w.van.velthoven@hum.leidenuniv.nl>

ROSS EMMETT
Arizona State University

Piketty’s history of equality expands the vision of his critique of inequality in previous work, most famously in *Capital in the Twenty-First Century* (2014), but also, *The Economics of Inequality* (2015). The impetus for the new book is to re-cast the arguments of his famous book—focused almost entirely on lessons learned from nineteenth and twentieth century economic history—into a longer historical narrative with more attention on the changes that are and could be happening, and who or what causes them (chapters 3–9). Not surprisingly, social class analysis is at the heart of his account, and the call for democratic socialist alternatives is more refined. While the narrative is briefer in the new book than it was before, the timeframe he considers is much longer, and the policy proposals are aimed at issues arising from inequality in a variety of different cultural settings.

Readers familiar with the “hockey stick of economic history” (Our World in Data 2017)—the long pre-1800 history of societies with small elites living well while most of the population lived just above subsistence level, followed by a rapid increase over the past two hundred years in almost every metric of economic progress—will recognize in Piketty’s history a dependence on the same story turned toward a different prediction about the future. Until about 1800, well over 90% of humanity lived in desperate poverty. Life expectancy remained at about 30 years of age, almost no one was well-educated, and opportunities were limited to a radius of about 30 miles. The 1800s changed all that; or at least, changed the economic prospects for the average European or North American tremendously. Productivity and resource growth at astonishing rates enabled population growth both within the native population and via immigration. Europe’s population doubled between the French Revolution and Lenin’s arrival at the Finland Station. North American population growth exceeded that: from a population of close to ten million at the end of the American Revolution, the US population grew to over 100 million by the
First World War (the consequence of high birth rates with moderate early mortality until the Civil War, aided thereafter by declining death rates).

For Piketty, it is the history of the next 100 years (that is, 1920 to 2023) that has been a mixed blessing. Pre-1900 capitalism had produced wealth unimaginable to the vast majority of humans who lived before 1800, and spread its benefits widely across the current human population. Just not widely and deeply enough, Piketty thinks. Since the mid-twentieth century, inequality has grown despite economic growth and the growth of welfare states, and still trends upward, making the elites wealthier and, therefore, more resistant to reforms on either side of the government’s budget. The poor are better off than at any previous point in human history, but the redistributive policies that aided the rise and increase of the upper and middle classes have not extended as many benefits farther down the income tables. The poor have therefore fallen behind, and the gap between them and the middle class has grown. For Piketty, what is worse is the erosion of the welfare state provisions that has occurred in the 21st century, which is extending the distance between the middle classes and the upper class.

We are facing a crisis point, Piketty believes. Either we begin a long retreat to the social inequalities of the early 1800s, with a status-based elite trumping a land-based one, or we move forward by accepting the primacy of political, economic, and social equality. The final chapter lays out Piketty’s views of how societies around the world could become more democratic, ecological, and multicultural; what he calls:

A democratic and federal socialism, decentralized and participatory, ecological and multicultural, based on the extension of the welfare state and progressive taxation, the battle against discrimination, educational equality, the carbon card, the gradual decommodification of the economy, guaranteed employment and an inheritance for all, the drastic reduction of monetary inequalities, and finally, an electoral and media system that cannot be controlled by money. These are only a few of the options. (237)

Rather than retell his narrative, which has been retold via academic presentations and media stories for the past several years, I intend to proceed directly to a criticism of Piketty’s argument built upon the work of Vincent and Elinor Ostrom.
The Ostromian response is one that Piketty goes out of his way to avoid. Elinor Ostrom is only mentioned in one footnote, and in that note he focuses solely on her notion of property as a “bundle of rights” (footnote 4, 34). Piketty includes the reference to her work simply as a means to lay the groundwork for his understanding of property, which includes real property, tangible occupation-related assets, and financial assets, minus debt (figure 4, 31). But Elinor’s construction of the means by which we may achieve long-surviving commons is problematic for Piketty’s top-down approach. Much of the human capital involved in commons’ management is embedded locally in a community whose aims will probably neither align with national nor international agencies, and their knowledge is simply better than any national (not to mention international) agency regarding the multiple factors that enable success in their own commons. Contra Piketty, then, Ostrom believes the local community’s discussion and decision-making should take precedent over any national or international decision-making, a focus made clear by her emphasis on the fallibility of humans and the necessity of humility regarding our knowledge and capabilities. She also recognized that governmental units will probably have goals at odds with what a community’s aims might be, setting up a conflict between jurisdictions (Ostrom 1990). In response, Piketty appears to argue that collating studies of commons around the world would enable principles of successful operation to emerge. But here Vincent Ostrom’s work further undercuts Piketty’s argument, by undermining his optimism about positive change coming from public officials. Vincent argued that the “omnicompetent public officials, omniscient legislators, and the perfection of a uniform and universal system of laws”—things Piketty assumes will guide us in the future—run contrary to human nature and assume “impossible conditions” (Fotos 2015, 68). The Ostroms’ work, therefore, suggests that Piketty’s (243) recipe for a universalistic sovereignty “contain[s] much bad medicine” (V. Ostrom 2008, 4).

A common theme in the work of both Elinor and Vince Ostrom is that experiments which run against the established patterns of life in a community are unlikely to succeed, and may well diminish a community’s long-term capacity to sustain itself. Elinor’s commons’ governance depends upon the historical development of patterns of interaction that are well-known within the community. Indeed, the history of every successful commons can identify key points where issues were settled by assent of its members, followed by adjustments that account for new issues or
threats arising within the community, the physical environment, or from higher-level governance structures. And there are plenty of stories regarding well-functioning commons (or efforts to restore a commons) that were undermined by a mandate from municipal, state, or national programs that claimed to know the ‘right way’ to do things, and then failed. What is often lost along with the commons is its community-minded orientation.

Thus, the Ostroms’ work ran directly counter to Piketty’s orientation. For him, moving governance upward is the logical thing to do because it enables a global cadre of economic experts to use generalized knowledge to address what look like issues common to many societies, so that common solutions can be created. The imposition of those common solutions would be aided, Piketty argues, by some common governance structures that would address similar issues in similar ways across the world. The question that concerns me is whether Piketty’s world is one that I would want to live in.

While reading A Brief History of Equality, I was also preparing to teach a course on Philosophy, Politics and Economics at my university (Arizona State). In the process of selecting materials for the students to read, I came across a Kurt Vonnegut short story that I had read long ago, and which I had thought outrageous (and a bit silly) at the time. The story is Harrison Bergeron (1961) which tells the tale of a man living in a society which is dedicated to making all people equal. In order to do so, the society’s ‘Handicapper General’ forces citizens to wear ‘handicaps’ that would diminish their strengths and enhance their weaknesses. Piketty’s objective seems to be analogous: to achieve equality across societies by enhancing or handicapping aspects of each society’s governance structure until no country is significantly different from any other.

The questions I would ask regarding Piketty’s proposals in his Brief History of Equality are similar to those suggested by Vonnegut’s story: What constitutes equality? Who decides? And, to what lengths are we prepared to go in forfeiture of freedom to achieve equality?

REFERENCES

Contact e-mail: <ross.emmett@asu.edu>

**Ahmed Al-Juhany**  
*University of Calgary*

Plenty of introductory books have explored the concepts and methods of evolutionary biology and economics (e.g., Sterelny & Griffiths 2012; Reiss 2013), but few have explored how those concepts and methods are shared and traded between the two disciplines. *From Evolutionary Biology to Economics and Back* helps fill this gap. The book’s project, as one of its authors puts it, is to present a systematic study of the “economics/evolutionary biology interplay” that’s marked much of the two disciplines’ histories (2). All those oft-cited moments—from Darwin’s inspired borrowings of Malthusian ideas to the interdisciplinary construction of evolutionary game theory—seemed to suggest deep similarities between the biological and economic domains and the ways in which researchers conceptualized each. Thanks to André et al.’s contribution, scholars now have an accessible introduction through which they can start to make sense of those similarities.

The authors anchor their study to a set of key concepts that both evolutionary biology and economics seem to share. The concepts are shared either in the sense of being picked out by the same linguistic term or in the sense of being made to serve similar functions in each discipline (2). ‘Competition’ is an example of concepts shared in the former sense. While in evolutionary biology it might refer to organisms’ relative degree of success in securing finite opportunities for survival and reproduction (48), in economics it can refer to rivalrous situations in which different agents try to trade with the same market participants (50). ‘Fitness’ and ‘utility’ are examples of concepts shared in the latter sense. Though distinct, the two concepts play similar roles in predicting things like phenotypes and choices when subjected to some "maximization principle" (2).

A lexicon-like list of twenty-five of these concepts makes up the bulk of the book (chapter 3). Each entry presents the same concept twice. First, one of the book’s authors describes the concept as it appears in its biological context, and then another describes it as it appears in its economic
context. After that, the two presentations are ‘synthesized’: their respective contents are compared and contrasted. The collection of syntheses emerging out of these presentations gives us a sense of how evolutionary biology and economics overlap and diverge conceptually.

The entry on ‘altruism’ can help illustrate how these concepts are presented (25–30). The concept is first described as it appears in evolutionary biology. There, it refers to the expression of characteristics that diminish an acting individual’s chances of reproductive success, their ‘fitness’ (93), but improve those of another. Its paradoxical nature—how could natural selection ever favor altruism if it reduces the fitness of the individuals expressing it?—and its theoretical explanation—it can favor altruism when the costs incurred by altruistic individuals are offset by the benefits conferred to their genetic relatives—are discussed. Clarifications and elaborations then follow.

The concept is then described as it appears in economics. ‘Altruism’, there, refers to the expression of ‘social preferences’ that preserve or improve other people’s well-being at the expense of an acting agent (28). The concept is important because it captures experimental subjects’ tendencies to deviate from the narrow, traditional conception of rational agents as self-interested egoists. This deviation has motivated a rich literature that models and tests hypotheses about the different patterns with which people express social preferences and behave altruistically (28–29).

The two presentations are then synthesized (29–30). The biological and economic concepts of altruism are shown to be similar in capturing self-sacrificial behavior that benefits others at the expense of an acting individual. But the kinds of sacrifices made and benefits conferred differ in each case. In evolutionary biology they are interpreted in terms of fitness, while in economics they are interpreted in terms of well-being. The concepts also raise different investigative questions in each discipline. In evolutionary biology the most salient questions are those concerned with the origin and evolution of altruism; in economics they are those concerned with how and when agents’ preferences deviate from those of a rational egoist.

The twenty-five entries do an excellent job introducing readers to complicated concepts. Whenever relevant, the authors disentangle common sources of confusion and point out prevailing disagreements. In the case of ‘adaptation’, for instance, the entry quickly distinguishes between four different concepts the term picks out, which are all too frequently conflated with one another (20–22). In the case of ‘information’, the entry
warns readers that, though popular, biologists’ metaphorical talk of genes ‘coding’ for or otherwise carrying information about phenotypes has been criticized for being overly simplistic and misleading (105). With careful clarifications and qualifications like these, readers are assured they are in reliable hands as they are guided through the complex, conceptual terrain lying at the intersections of evolutionary biology and economics.

But how are the entries related to one another? In chapters 2 and 4—both written by Bernard Walliser—they are said to be related by way of analogies. Not only do many of the concepts themselves resemble one another in certain respects, as we have seen, but researchers have historically recognized the concepts’ resemblances and have sought to leverage them in their investigations. Those researchers would note similarities in how certain biological and economic phenomena were conceptualized and would infer, heuristically, that some other biological and economic phenomena might be conceptualized similarly too. This sort of analogy-making is clearest in the development of ‘universal Darwinism’ (Dawkins 1983; Dennett 1995)—the view that entities in many non-biological domains, economics included, undergo the same processes of variation, selection, and inheritance as entities in the biological domain. It is on the basis of such an analogy that some researchers have inferred that firms and their routines evolve through a Darwinian process of social selection (e.g., Hodgson and Knudsen 2004, 2010).

In the first chapter on analogies (chapter 2), Walliser introduces a three-piece typology meant to capture the conceptual similarities between evolutionary biology and economics (8–14). The strongest analogies are ‘integral’. They are grounded in the claim that models not only have “equivalent structure[s]” and “comparable interpretation[s]” but also apply to the same material domain (e.g., the biological domain) (8). ‘Substantial’ analogies are weaker. They are the sort of analogies grounded in the claim that models have equivalent structures and a likeness in interpretation, though they apply to distinct domains (e.g., the biological and economic domains). The weakest analogies are purely ‘formal’. They are grounded only in the claim that models have equivalent structures, regardless of how they are interpreted or where they apply.

Walliser then uses this typology to examine how the concepts in the entries might be brought together to better understand the analogous ways in which biologists and economists conceptualize phenomena (chapter 4). He considers for instance how they conceptualize behavior—
of non-humans in the case of biologists, of humans in the case of economists. Both depict their entities as maximizing some ‘objective-function’ in the presence of some constraint (169). Here, the concept of ‘optimization’ (120–126), which essentially refers to this pattern of maximization, helps us capture an important formal analogy: both disciplines conceptualize their entities as exhibiting optimizing behavior.

Walliser also shows that we can make this analogy substantial if we consider the processes by which non-humans and humans come to behave optimally. Both non-humans and humans might be interpreted as engaging in comparable but ultimately distinct processes of deliberation. Walliser consults another concept—that of ‘trade-offs’, which designates an “area of possible combinations [of alternatives] and the solution” in this area that will be selected by the optimization procedure (58)—to clarify the similarities and differences between them. Though both biologists and economists might interpret their target entities as engaging in some kind of deliberation about trade-offs, Walliser tells us, biologists take their entities to be performing “mechanical responses to external stimuli” while economists take theirs to be making calculations on the basis of beliefs and preferences (169). With the concept of trade-offs in hand, we can make out a substantial analogy between the deliberations taking place in the separate domains of biology and economics.

Walliser’s application of this typology shows it to be a useful tool for classifying the analogous conceptual spaces of evolutionary biology and economics. But his use of it, at times, can be contentious. For example, he argues that “no integral analogy can be stated between biology and social science fields like economics because of the emergence of ‘thought’ during the evolution process” (171). This implies that humans and their economic affairs are just far too distinct to be conceptualized as belonging to the same domain as purely biological phenomena. As for universal Darwinism’s famous attempts at establishing integral analogies, Walliser writes that such attempts can only be “sustained [by making a] gross generalization which erases all the details” of relevant phenomena (15). Analogies lose their usefulness in cases like these because “they no longer [allow] the transfer of specific and testable properties” across contexts (15).

This is a harsh indictment of research programs that have been built around integral analogies. Hodgson and Knudsen’s (2010) brand of universal Darwinism—generalized Darwinism—is an obvious example of one such program. Like all universal Darwinists, Hodgson and Knudsen claim
that both biological and economic systems undergo processes of variation, selection, and inheritance. But there is more to their view than that. They argue that the fact that both kinds of systems undergo those processes implies that they belong to the same general domain: that of ‘complex population systems’ (26).

Walliser’s assessment of integral analogies would have us reject Hodgson and Knudsen’s program outright. But generalized Darwinism can be heuristically valuable (see also Reydon 2021). Some of its components have found productive use in fields like management studies (Becker 2019). Its focus on habits and routines as social ‘generative’ structures—i.e., as conditional mechanisms that turn environmental signals into ‘developmental instructions’ (Hodgson & Knudsen 2010, 113)—has especially helped researchers develop models of organizational routines (e.g., Pentland et al. 2012) and flesh out new ontological accounts of firms (e.g., Chassagnon 2014). Examples like these should give us pause: it is at least possible for integral analogies to be useful when researchers employ them creatively. We should not be so quick to dismiss them.

In the last chapter of the book (chapter 5), Johannes Martens concludes with an overview of the most recent interactions that have taken place between evolutionary biology and economics. He notes several trends: the mainstreaming of evolutionary game theory, the decline of evolutionary economics, and the rise of behavioral evolutionary economics. The historical and future progressions of those trends are sure to be treasure troves of insights for those looking to learn more about how biologists and economists leverage analogies, trade conceptual tools, and more broadly engage in interdisciplinary research.

On the whole, André et al. have done a wonderful job with this book. By mapping out key conceptual similarities and articulating useful notions of analogies, From Evolutionary Biology to Economics and Back equips readers to dive into the rich conceptual space that brings together biological and economic science. For new entrants looking to explore this space, I can think of no better introduction with which they might begin.

REFERENCES


Ahmed Al-Juhany is a PhD student in Philosophy at the University of Calgary. His research interests are in the general philosophy of science, philosophy of biology, and philosophy of economics. His dissertation investigates conceptual and ethical issues that arise in the biomedical sciences of aging.

Contact e-mail: <ahmed.aljuhany@ucalgary.ca>

WILLEM VAN DER DEIJL
Tilburg University

A MESSY STATE OF AFFAIRS

Science, and social science in particular, is often motivated by the promise of progress: making lives better, or improving human wellbeing. But seeing how central this motivation has been to science, it is perhaps surprising that the science of wellbeing is still in its early stages. This is not because of a lack of interest in wellbeing measurement by scientists (see Angner 2011 for an excellent historical overview). Rather, it appears to be because measuring wellbeing is hard. Developing measurements is always hard (see for instance Chang 2004 history of thermometers), but there may be something particularly difficult about wellbeing measurement. Wellbeing, in all plausibility, has a lot to do with qualitative assessments of mental states (at least in part). However, our access to the mental states of others is quite limited, especially if it comes to making comparative judgments of specific qualitative features.¹ ‘Is the color red you see, the same as the color red that I see?’ But also, ‘Does chocolate taste as good to you as it does to me?’ These questions are hard to answer. And measuring wellbeing seems to involve at least some questions in this category. But if this were it, things would not be so bad. After all, there are many psychological measures of mental states. What makes things particularly challenging is that wellbeing is a rather difficult concept to specify. What does it entail for a life to be lived well. This is, unfortunately, a rather philosophical question. And philosophers are divided on the subject. Broadly speaking, there are three strands of views: lives are good for those living them because 1) they are enjoyable (and are void of suffering), 2) contain high levels of desire-satisfaction (and little frustration of desires), and 3) involve numerous goods such as knowledge, achievement, and friendship, independent of whether they are pleasant or desired (while avoiding their counterparts).

¹ This is not to say that our access to our own mental states is all that much better (see Schwitzgebel 2008).
Overcoming these two challenges appears to require a multi-disciplinary effort. The question of what wellbeing entails is not one that can be empirically answered. And the question of which measure best captures wellbeing is not one that can be determined from the philosopher's lazy boy. There have been many different efforts trying to foster multi-disciplinary cooperation. There is a multidisciplinary *International Journal of Wellbeing*, founded by two philosophers, and initially empirically minded journals, like the *Journal of Happiness Studies*, have also published many contributions by philosophers. Still, genuine multi-disciplinary approaches have been rare. Empirical researchers publish studies using specific survey measures such as 'life satisfaction' or 'positive and negative affect' that have become widely accepted. They shun debates about whether these notions truly identify with how good life is for the individuals filling in the surveys, and contend that they are likely to approximate wellbeing to some extent. At the same time, philosophers have debated the nature of wellbeing without an emerging consensus. All and all, little headway appears eminent. At the same time, the need for policymakers to build on better measures than GDP to capture progress has not been resolved.

**Theory to the Rescue**

This is where Mark Fabian’s *A Theory of Subjective Wellbeing* comes in. Building on Anna Alexandrova and Daniel Haybron’s (2016) plausible diagnosis, Fabian starts off from the idea that progressing this debate further requires better theory. And this is what Fabian’s book offers: a theory. As the cover promises: Fabian’s theory “integrates not only ideas in SWB studies and analytical philosophy, but also ideas from clinical, moral, and developmental psychology; continental philosophy and welfare economics”. The reader is in for a hell of ride: from Nietzsche, Sartre, and Kierkegaard, to self-determination theory and the capability approach. It is a ride that is worth taking though. Fabian’s book is incredibly creative and insightful. Fabian diagnoses the state of research on wellbeing in different subfields as ‘disjunctivitis’, which is to say that “when researchers work in disciplinary silos and are disinclined to theorize they are prone to produce internally inconsistent bodies of knowledge” (43). Fabian describes the aim of the book, then, as follows: “my contention is that a
A unifying project is overdue in the field of SWB and I humbly present this book an attempt at it” (43).

THE THEORY
In order to explain the main gist of the book, we need to survey a number of key concepts that different fields of study tend to view differently. Wellbeing, or prudential good, is what is good for someone (philosophers typically add: for ‘the sake of the individual’, but Fabian does not). This notion is often used interchangeably with welfare in the philosophical and economic literature, but Fabian defines welfare as the standard of living. Subjective wellbeing (SWB) is the “individual’s own judgment and perceptions of their wellbeing” (4), and is also a serious contender for what wellbeing may be, substantively. Subjective well-being (note the hyphen!), however, is the research field that has been studying life satisfaction, and positive negative affect, and meaning in life. Fabian refers to this as a ‘school of thought’ sometimes, but sometimes as a body of research.

Fabian’s primary aim is to construct a theory of SWB. He does not commit to the view that SWB is also wellbeing, but “leave[s] that debate to the philosophers” (46). At the same time, he presents a fierce defense of his theory of SWB as an account of wellbeing, at least from commonly discussed problems (the experience machine, the immoral but well objection, adaptation, etc.). At the same time, he suggests that even if SWB indeed would be what wellbeing is, then still it should not be the aim of governments to maximize it. Instead, they should maximize welfare. Moreover, SWB can be measured, although most likely not in precise interpersonally comparable ways.

Most of the book develops the theory of SWB. Fabian expresses this theory in the form of an equation:

\[ \text{SWB} = f(\text{hedonia}, \text{eudaimonia}, \text{conscience}) \]

He calls this the SWB production function, analogous to production functions in economic theory. A chapter is dedicated to each of these three factors in the production function. The production function itself—the way these elements generate SWB—is described as the coalescence of being. Let’s look at these elements briefly.

Hedonia, here, refers to affective states. Someone does well here if they have an overall balance of positive over negative emotions. Eudaimonia is not so much understood in an Aristotelian sense, but rather in
the sense of Self-Determination Theory: being able to live according to your own values and pursue them successfully and autonomously. It may be surprising to philosophical readers to find eudaimonia as an element of SWB, but the reason for this is that in the intended sense, eudaimonia does not require one to live morally, but it requires one to live autonomously, according to one’s one intrinsic values. Finally, the incorporation of conscience is particularly interesting.

Here, Fabian joins up two literatures: the psychological literature on the importance of hope, having a sense of meaning, and purpose; and the existentialist philosophical literature. Fabian builds on Nietzsche and Kierkegaard to argue that existential dread is a common experience (in the West? I ask because Fabian is not so clear on this). Losing the sense that there is a divine standard of living can make us feel anguish, despair, and mostly, meaninglessness. All prudential and moral rules appear utterly arbitrary. And, indeed, psychological research suggests that this sense of despair, anguish, and arbitrariness tends to be utterly detrimental to one’s subjective wellbeing. Fabian contends—along with some existential thinkers—that there is no point in attempting to discover some objective standard. After all, psychological research suggests that moral intuitions have evolutionary explanations, debunking them as serious guides to moral truth. However, one can construct one’s own subjective standard. And doing so will give you a sense of meaning and purpose that you can take seriously.

The final element of Fabian’s theory is the coalescence of being: the idea that we can harmonize our actual self (how we are now), our ideal self (how we want to be), and our ought self (how we feel we should be). Feeling like you need to be married by 30, wanting to pursue a career in philosophy, and currently being 29 without prospects of marriage and a career in philosophy can clearly make you feel bad about your life. Resolving this conflict, then, for instance, by getting rid of the idea that you need to get married, and starting a PhD in philosophy instead, is one way to find a coalescence of being.

So much for the theory. Does it succeed in its aims to develop a guiding theory for wellbeing research?

There is a lot to say here. Fabian is perhaps right that the analytic philosopher’s need for clarity sometimes gets in the way of making progress. Still, there is one aspect Fabian deliberately undertheorizes that worries me. A question that easily comes up when we look at the SWB
production function that Fabian postulates is whether the different elements (hedonia, eudaimonia, and conscience) are supposed to be constituents or causes of SWB. Fabian thinks that this is not a very meaningful question:

[the] debate over what is “intrinsic” to well-being and what is merely “instrumental” is counterproductive. Mental states, preference satisfaction, nature fulfillment, and process are all integral to the model. (7)

But this strikes me as problematic. Take hedonia. It seems fairly obvious that not feeling well is bad for someone, and that feeling well is good for someone. Even those who are not much inclined to hedonism take that to be true (see for instance Bradford 2021 for a discussion). However, the same is not obviously true for relatedness, or acting autonomously (or having a sense thereof). Often lacking these goods will not be good for someone, but there are plenty of instances to think of where that does not appear to be the case (I will happily feel unrelatedness in a group of extreme right-wing protesters). In other words, not all the elements of Fabian’s production function appear to be equal. If all these elements would be constituents of SWB, then this raises a complicated issue: how important are all these elements relative to each other? In other words, how do we get to know all the numerical values in the function? But it is not at all clear that all these variables are constituents.

To clarify, let’s build on Fabian’s analogy and imagine a function describing national GDP growth that would look like this:

\[
\text{GDP growth} = f(\text{quality of economic policy}; \text{technological innovation}; \text{GDP growth of the central province})
\]

A function like this may very well be a good proxy for GDP growth. However, it mixes in both causes (quality of economic policy and technological innovation) with constituents (GDP growth of central province). This may be fine for developing proxies, but it would look bad in an economic theory.

Fabian’s suggestion that this is not a very important issue is, I think, mistaken. Imagine that a group of individuals are suffering a bout of anhedonia. Treating anhedonia is difficult, but imagine that we can make the individuals feel like they have more autonomy over their life. Even
though none of the individuals will be able to feel positive emotions, they are now better able to act in accordance with their aims. It seems that if this does not make them feel better, this is little consolation. Perhaps Fabian will point out that this does not matter much in practice. These cases may be rare. When discussing the problem of immoral people who are still high on SWB, Fabian suggests that he takes these individuals to be exceptions. And he is fine if his theory only covers the general population. But there is an important reason that philosophers like to argue with counterexamples. Imagine that the response to Einstein’s discovery of relativity would have been: ‘this is an exceptional case, and we are fine if our theory of physics just covers the non-exceptional circumstances’. This would have been an obvious mistake, and we would have forgone a significant opportunity to learn about the laws of thermodynamics. This is why exceptions matter, especially for theorizing.

Moreover, the truth does matter in practice. The identification of wellbeing with preference-satisfaction has, according to many critics, led to an overemphasis on making people financially autonomous, without focusing on how people feel. If anything, wellbeing theory plays exactly this practical role: determining which aims deserve particular attention. Casting this aside as philosophical frivolity may not do justice to the importance of determining what exactly it is that makes our life good for us.

Another issue is the conception of SWB at the heart of the book. For a book that is called *A Theory of Subjective Wellbeing*, the book spends surprisingly few words discussing the nature of the concept of subjective wellbeing. It is how well an individual is doing according to their own standard. But there are more questions to be answered. Take a key issue: whether or not SWB is a mental state. Fabian writes: “I am even reluctant to describe SWB as a mental state because the self is embodied, and coalescence speaks to harmony between thought and action” (189). This is surprising considering the way that SWB is typically seen (as, indeed, a mental state). Later, discussing the benefits of SWB as a theory of wellbeing, it becomes clear that Fabian does see the concept as a mental state (for instance, in chapter 11, Fabian bites the bullet on the experience machine objection).

But also consider the issue of whether SWB can be compared across individuals. There is an excellent section on measurement (chapter 10) that discusses how the measurement of SWB faces the problem of scale norming: the same level of SWB can be judged differently in different con-
texts. Someone who has not lived a very good life, but then starts to improve can rate their life 7 twice, but because their perspective has shifted, that 7 might mean different things. This is indeed a difficult problem for comparing measures of SWB (see for instance Ingelström and van der Deijl 2021), however, if SWB just is how well someone is doing according to their own judgment, you may wonder if it is even conceptually possible to compare across individuals and time. Scale norming, then, is only a symptom of a deeper problem: that the concept does not allow for interpersonal comparisons. A judgment of ‘I am doing quite well’ of one person may reflect a less enjoyable life than someone else who would say ‘I am not doing very well’, but enjoyment is not the standard of wellbeing. Individual judgments are. We may say that person A is doing better according to their standard than B is doing according to theirs, but this does not tell us much about which life is better for them, if these standards are different. Can SWB ever escape the problem of scale norming, even outside of the context of measurement?

That being said, it takes intellectual courage to write a book touching on so many different research fields. I am sure that Sartre scholars (of which I am not) would find some objections to the way that Fabian discusses him, and I am sure that analytical moral philosophers will find the conclusion that no objective moral standards exist because our moral intuitions have evolutionary bases a little quick, and I am sure empirical SWB researchers may find flaws that I do not see. However, this is merely the result of undertaking such a courageous task. Fabian is right to say that the field of wellbeing research needs better theories, and A Theory of Subjective Wellbeing is an excellent step in the right direction.

REFERENCES
Willem van der Deijl is assistant professor of philosophy at Tilburg University. He was an editor of the Erasmus Journal for Philosophy and Economics from 2014-2017. Contact e-mail: <wjavanderdeijl@gmail.com>

LENNART B. ACKERMANS
Erasmus University Rotterdam

Thomas Kelly’s *Bias* is a wide-ranging philosophical study of the concept of ‘bias’ in the pejorative sense as well as (in one chapter) the non-pejorative sense. The English word ‘bias’ is used in many different ways to indicate that a person, procedure, or outcome is flawed. The book is concerned with all of these ways. It is an ambitious project looking to give a general theory that explains what bias is and how its different uses relate to each other. It also explores some independent epistemological questions related to the phenomenon of bias.

The book is divided into three parts. Part 1 is called *Conceptual Fundamentals*. It explores some basic questions about the concept of bias, such as to which types of things the label applies and whether some of them are more fundamental. Among other things, bias can be attributed to people (judges, scientists), mental states (reasoning processes), procedures, and outcomes. Kelly argues that there is no fundamental hierarchy among these different types of bias, although sometimes a bias in one sense exists because of bias in another sense—such as when a person is biased because their reasoning process is biased.

Part 2 is called *Bias and Norms*, and it lays out Kelly’s theory of bias called the *norm-theoretic account of bias*. In a nutshell, the theory holds that a bias involves a systematic departure from a genuine norm or standard of correctness. There is a great variety of norms that Kelly considers genuine norms: among them are ethical norms, epistemic norms, and norms of rationality. People or things are biased when they depart from such norms systematically, that is, consistently in a particular direction.

For example, suppose a judge generally rules in accordance with the evidence. When the defendant is black, however, non-evidential considerations—namely, prejudice—affect how she comes to her legal decisions. According to the norm-theoretic account of bias, this judge is biased in multiple ways. She is biased with respect to the epistemic norm of believing in accordance with the evidence since her departures from this norm happen systematically (when the defendant is black). She is also biased with
respect to a norm of justice according to which only the guilty should be punished since she punishes innocent black people more often than innocent non-black people. These are just two of the many ways in which one might consider this judge biased on the norm-theoretic account.

The norm-theoretic account allows Kelly to draw some plausible conclusions that appear interesting. For example, Kelly gives a novel explanation of the bias blind spot—our tendency to see bias in other people in ways that we fail to see in ourselves (see also Kelly 2024). The relevant norm, according to Kelly, is a norm of accuracy: people ought to believe things that are true. One is biased with respect to accuracy, call it truth-biased, when one's beliefs systematically depart from the truth. Kelly's insight is this: given that we believe as we do, a person who has systematically different beliefs than our own appears to depart from the truth systematically. Hence, in light of our own beliefs, we ought to think this person is truth-biased. At the same time, we ought not to think of ourselves as being truth-biased.

Part 3 of the book is called Bias and Knowledge, and it is the most interesting part of the book. It explores epistemological questions related to bias that are to some extent independent of the theory of bias expounded in part 2. Some of its results are: (1) one can genuinely know something even if one is biased with respect to that knowledge; yet (2) a token belief that is a manifestation of bias is (probably) not knowledge; (3) biased people can sometimes be reliable; and (4) it is sometimes rational for laypeople to dismiss the views of experts because they think they may be biased. In what follows I discuss my two main criticisms of the book.

DOES THE CONCEPT OF ‘BIAS’ HAVE A NATURE?

The word ‘bias’ is used in many different ways. A wide variety of things can be biased, and they can be biased in multiple senses. Kelly thinks the explanation of this fact is not linguistic. Rather, the diversity of types of bias “reflects something deep about the nature of bias itself” (9). Parts 1 and 2 of the book are devoted to finding out what this nature is—or at least to learning more about the nature of bias, by answering questions such as: What kind of things can be biased? Is there a hierarchical structure between types of bias? I have trouble making sense of this project. Finding out the nature of something presupposes that it has a nature. Does ‘bias’ have a nature?
First, let me clarify the opposing viewpoint which I find more attractive. The word ‘bias’ means what we want it to mean. The reason that English has the word is that it has apparent usefulness in the English-speaking world. At any point, we might decide to change its meaning because we find that more useful (or it might evolve naturally). Nothing is lost by such a change, because ‘bias’ does not refer to something fundamental that exists independently of culture and language, as Kelly suggests. That is not to say that instantiations of bias don’t exist, such as biased judges. My position is that the overarching concept of bias does not refer to an existing thing, except if what exists is the linguistic reality or the definition chosen by humans.

On this view, nothing is surprising about the fact that so many different things are called biased, in so many ways. That is simply a result of the fact that the word bias means many different things. We would have too many words if each of these meanings was given a separate label. The similarity between the different meanings, combined with the etymological history of the word ‘bias’, explains that we have a single word for all of these things. Hence, no theory of bias is needed to explain how the different uses of ‘bias’ relate to each other.

Kelly suggests that my kind of story is implausible:

The fact that we sometimes use the word ‘biased’ to convey a negative evaluation of the object of our attribution, but in other cases, we don’t, seems importantly different from the fact that we sometimes use the word ‘bank’ to talk about financial institutions and sometimes use the same word to talk about riverbanks. (147)

Here and in other places, Kelly suggests that he thinks a linguistic explanation that the word ‘bias’ has different meanings is implausible and that we should therefore look for an explanation referencing the nature of bias.

I find some evidence for my position in the fact that not all languages have a similar word. My own native language, Dutch, has no direct translation for ‘bias’: it is translated in different ways depending on context. In Kelly’s view, this should mean that Dutch is missing an important concept, such as we might say when a language has no word for ‘cloud’, ‘truth’, or the number ‘5’. However, I don’t think Dutch is missing anything at all. We are doing fine with different words.
Unfortunately, Kelly does not give many reasons for believing that ‘bias’ has a nature. He does give a comparison of his project to the philosophical project of understanding the nature of truth (43–44). About truth, one can ask: which types of things can be true or false? Kelly maintains he is asking such a ‘nature-of’ question about bias: which types of things can be biased? At least in the case of truth, many philosophers would agree that this question can be understood as a nature-of question, rather than a linguistic question (‘To which types of things does the English language attribute truth and falsity?’).

The book is missing an argument that such nature-of questions about bias make sense. Concepts like ‘truth’ and ‘5’ can reasonably be thought to exist independently of human definitions and conventions—and even in those cases, there are philosophers who disagree. However, I don’t know any good reasons to think that the same is true for the concept of ‘bias’.

While this is a major problem with part 2 of the book, the norm-theoretic account of bias is still useful, although perhaps in different ways than the author intended. First, it is useful as a theory for understanding the linguistic practices of attributing bias in English. Second, the concept developed by the theory might be useful in the sense of conceptual engineering. The concept of ‘bias’, as characterised by Kelly, might play a useful role in society, science, and philosophical subjects like social epistemology.

**IS THERE A GENUINE EPISTEMIC NORM OF ACCURACY?**

In the more epistemological sections of the book, Kelly gives much importance to one specific type of bias that I called truth bias. It plays a central role in his explanation of the bias blind spot (chapter 4), his argument that people should rationally attribute biases to others (chapter 3), his chapter on the epistemology of disagreement (chapter 10), and other chapters. It also features prominently in a recent paper (Kelly 2024). These discussions rely on the view that there is a genuine norm of accuracy, which requires people to have beliefs that are true. I have some reservations that I explain below. First, I am sceptical that there is such a norm. Second, even though truth bias might be a legitimate bias, I think

---

1 For an overview of positions that truth does not have a nature (deflationism), see Armour-Garb, Stoljar, and Woodbridge (2023). For an example of a position that mathematical objects have no independent existence (nominalism), see Azzouni (2004).
the book overstates its significance. Other biases and epistemic norms may be more important.

To be clear, it is certainly desirable that our beliefs are true. This widely shared desire explains why we have many of the epistemic norms that we do. However, as a matter of logic, from ‘it is desirable that one believes $X$ if and only if $X$ is true’ it doesn’t follow that ‘one ought to believe $X$ if and only if $X$ is true’.

The norm that we ought to believe what is true does not seem to be a genuine epistemic norm to me. This becomes clear if we consider cases in which the truth norm is violated but other epistemic norms are not. Suppose that I put some ice cubes in a jug of water on the kitchen counter. Three hours later, without returning to look at the jug, I form a belief that the ice cubes have melted. (All this time the room temperature stayed at 25 degrees Celsius.) However, it is a well-known fact of statistical mechanics that the ice cubes might still be there, although the chances of this happening are extremely small. According to the above accuracy norm, I ought not to believe that the ice cubes have melted in a possible world in which they have not. However, it seems I ought to believe this in every world in which I have the evidence that I do, including the extremely unlikely worlds in which my belief is false.

The accuracy norm prescribes that the beliefs of people should be dependent on facts that they are unaware of and are in no position to be aware of. Hence, the norm seems to violate an epistemic version of the commonly accepted ethical principle ‘ought implies can.’ In the ice cube case, I would be required to change my beliefs depending on whether an extremely unlikely event happened that I can't possibly know about or be required to know about. This strikes me as implausible.

Although there seems to be no epistemic norm of accuracy, there is a way out for Kelly. Instead of a norm, there could be a standard of correctness that beliefs should be accurate. After all, biases, according to the norm-theoretic account, involve systematic deviations from genuine norms or standards of correctness. Hence, Kelly could maintain that truth biases are biases in this second sense. This seems alright to me. However, while there may be truth biases in this sense, I doubt that they should be given as much significance as Kelly does.

Consider one of his own examples (75-76). Two persons, Left and Right, disagree on politics in a systematic way. Left thinks that Right’s positions tend to be too much to the right; Right thinks that Left’s positions tend to be too much to the Left. Kelly shows, convincingly, that both
ought to think that the other person is truth-biased, on pain of irrationality. Kelly presents this as an interesting result. But why is this any more interesting than the platitude, ‘I believe I am right, so I ought to believe you are wrong’? A judgement of truth bias could be interesting if it had implications, such as epistemic implications about which substantive propositions one ought to believe. In the situation described here, I don’t think there are any such implications.

Things become admittedly interesting when there is evidence that a person or group of people are truth-biased with respect to some topic. Kelly introduces an example in which scientific studies show that people like Right often have inaccurate opinions. Kelly argues that Left, in such a case, has reason to discount Right’s opinions. This conclusion seems plausible.

In other cases, however, Kelly goes too far with his claims that we can discount others because they are truth-biased. In one example, we imagine that Antonin Scalia’s views on the constitutional legality of restrictions on abortion were insensitive to the truth (suppose we are told this by an infallible oracle). That is, Scalia, who believed abortion restrictions are constitutional, would have also believed that abortion restrictions are constitutional if they were not. Curiously, Kelly claims that Scalia’s opinion in this situation is “evidentially worthless” (216). This strikes me as incorrect.

Kelly’s claim seems to be that Scalia’s view is evidentially worthless only because the view is truth-insensitive. This would imply that Scalia’s truth-insensitive views are evidentially worthless even if Scalia does not violate any other epistemic norms, including norms about responsiveness to evidence. But in that case, the ice cube example is similar. Since I don’t violate other epistemic norms (we assumed), my beliefs, while insensitive to the truth, are properly sensitive to the evidence. For example, if I received additional evidence that my clock malfunctioned and instead only 10 minutes had passed, I would change my beliefs about the ice cubes. On the other hand, my belief that the ice cubes have melted is insensitive to the truth, holding fixed my evidence. However, it is clear that my beliefs have evidential value. My telling you that the ice cubes have melted should increase your own belief that they have.

Turning back to the constitutionality of abortion, suppose similarly that Scalia does not violate epistemic norms besides the (disputed) norm of accuracy. Thus, the infallible oracle’s pronouncement must mean that
Scalia’s views would be the same whether or not abortion is constitutional, holding fixed his evidence. However, possible worlds in which abortion is constitutional while Scalia’s evidence is the same might be very unlikely, for all we know. In fact, that is what we ought to believe if Scalia is an expert (assuming we have not heard contradictory opinions from other experts). Hence, just as in the ice cube example, Scalia’s truth-insensitive view has evidential value.

Another problem for the significance of truth-bias is that we typically don’t know that we have truth biases. Rational and justifiable beliefs are often false. Since we are normally not able to know that they are false, there is not much we can do about that. On the other hand, there is a lot we can do about biases related to norms of evidence, such as biases with respect to the norm of having justifiable (rather than true) beliefs. It would have been interesting to see more discussion on that front.

Despite its shortcomings, *Bias* is an impressive work with many interesting discussions on a broad range of topics. While I would not recommend reading it in full, scholars of bias would do well to check what Kelly has to say about this or that topic, particularly in the epistemological part of the book.

**REFERENCES**


**Lennart B. Ackermans** is a recent PhD graduate of Erasmus University Rotterdam.

Contact e-mail: <philosophy@ackermans.ch>
PHD THESIS SUMMARY:
The Singular Plurality of Social Goods:
Social Ontology and Collective Dilemmas

MARCO EMILIO
PhD in Human and Social Sciences - Philosophy, April 2022
Université de Neuchâtel - Institute de Philosophie

Today, the ontological status of economic entities has become increasingly significant for different branches of philosophy (see Lawson 2003). Among such entities, public goods have attracted special attention in metaphysics due to their crucial role in shaping institutional actions. However, recent efforts to cope with global challenges (explicitly or implicitly) based on the theory of public goods appear incapable of fostering sufficient cooperation, as evidenced in the case of coping with climate change. Moreover, from a theoretical angle, it appears to be difficult to reduce such goods to the pursuit of individual goods. The sum of individual selfish acts is simply insufficient for explaining the emergence of the complex coordination necessary for supplying collective goods in large groups (Olson 1971), as evidenced in cases of common pool resources management (Ostrom 1990) and political participation (Uhlaner 1989).

These shortcomings suggest the relevance of an inquiry into the different standard types of public and private goods as well as the classic categorization criteria of excludability and rivalry. Such an inquiry raises two key questions: (i) what kinds of goods are at stake in economic processes; and (ii) how can a more thorough theoretical image improve the understanding of institutional actions and collective responsibilities?

In this thesis, I address this matter by proposing the possibility of a coherent ontological taxonomy of irreducible and reducible social goods. By drawing on current research on collective intentionality within social ontology, I investigate the interplay between the categorization of a good, its robust provision, and its ontological dependence on socio-relational entities. By framing collective action gridlocks1 through the lens of this suggested taxonomy, I re-examine prevailing approaches to social

---

1 According to influential global policy authors, when a “specific set of conditions and mechanisms” (Hale, Held, and Young [2013] 2014, 3) impedes the demanded international cooperation, there is a gridlock.
dilemmas influenced by the traditional economic categorization of goods and instead argue for the consistency of a pluralistic account of social goods.

This dissertation consists of three main sections. In the first section, I investigate how the standard categorization of public and private goods in economic theories (Samuelson 1954, 1955) has influenced the ethical and political landscape of current collectivities. By drawing on disciplinary literature, I specifically examine how the standard categorization of goods has hampered global cooperation in at least three different areas: climate change policies (Hale, Held, and Young [2013] 2014), the fight against disinformation (European Commission 2018), and efforts to improve the quality of educational services (Anomaly 2018). I then explain how this shortcoming seems strongly related to two theoretical tendencies: the ontological bracketing of socio-relational entities (legitimacy and consensus, for instance) and a reductive understanding of the genesis and persistence of coordination involved in supplying and enjoying economic goods (Pecorino 2015).

Furthermore, I examine proposals based on the notions of common-pool resources (Ostrom 1990) and relational goods (Uhlaner 1989), suggesting that they have been unable to cope with the drawbacks of the standard categorization of economic goods due to their conceptual dependence on rational choice theory (Herfeld 2018) and methodological individualism. The first section ends with an examination of how the current research on collective intentionality seems to provide a valuable alternative for inquiring into the ontological categorization of a good as either private or public.

The second section develops a refutation of the ontological reduction of all economic and social goods to the level of the individual, arguing instead that a form of cooperation grounded in we-intentions and we-reasoning attests to the plausibility of certain irreducible collective goods. This argument is accompanied by the presentation of a comprehensive taxonomy of goods. It begins by assessing Charles Taylor’s views about irreducible social goods based upon some form of shared background, highlighting the need to clarify how individual attitudes combine into a “common understanding” (Taylor 1995, 139). By inquiring into the interdependence between climate change mitigation and environmentalist culture, I argue that a global public good, such as mitigation (Kaul 2016), depends on some form of shared background between individuals. It then presents the crossroad problem (Schweikard and Schmid 2021) of the
whole thesis as a coupling of the two following conflicting statements: an irreducible social good is not an aggregate, a sum, or a collection of individual goods (irreducibility thesis); nonetheless, individuals recognize irreducible social goods through individual evaluative acts or judgment (individualistic ontology thesis).

With this problem in mind, I dissect the excludability and rivalry criteria and attempt to show how they are ontologically dependent on individuals’ intentional attitudes, judgments, and actions. Following this line of thought, I draw on Searle’s theses (2010) to suggest that the existence of a social good necessarily depends on cooperation as a full-blown form of collective intentionality. By drawing from empirical literature (Garnett, Ealy, and Lewis 2015), I provide a sketch of different forms of cooperation, in order to show that achieving a common aim can involve various forms of belief sharing and coordination between collaborative actors. Following Tuomela (2013), three forms of cooperation are outlined: I-mode cooperation, pro-group I-mode cooperation, and cooperation in we-mode. The section then analyzes the issue of the ontological robustness of cooperative processes, with a particular focus on the observation that human collaboration is often dynamic and multi-layered (and can evolve or fade) (Pettit 2015). By relating different types of cooperation and categories of goods, I outline a three-fold hypothesis of a taxonomy of social goods: private social goods that depend on I-mode cooperation; public social goods that rely on pro-group I-mode cooperation; and collective goods based on we-mode cooperation. This is followed by an examination of the interdependence between the different types of goods along the axis of ontological robustness.

Moreover, I argue for the existence of cooperative relational goods by investigating how peculiar relational entities (such as trust) seem to be a necessary precondition for the durability of cooperative forms. On the other hand, the evolution of specific forms of cooperation can be linked to the variation of payoff sharing and expectations regarding modes of coordination. Therefore, by recalling the construct of collective sensemaking (Weick, Sutcliffe, and Obstfeld 2005), I propose that specific forms of reasoning can facilitate the evolution from I-mode to pro-group I-mode and, eventually, to we-mode cooperation. Consequently, shared reasoning practices can become the object of intentional enhancement and thus be categorized as a developmental cooperative good.

Finally, in the third section, by applying all these hypotheses to the cases of action gridlock presented in the first section, I provide a test of
the consistency of my proposed taxonomy. This part first results in a con-
ceptual revision of the notion of rivalry as an emergent by-product of
subjects’ intentional attitudes towards the same target. It also investi-
gates the emergence and persistence of cooperation in uncertain environ-
ments, stressing how reasoning transformation and fairness are crucial in,
for instance, common-pool resources management. Moreover, the in-
terplay between developmental cooperative goods and relational ones is
highlighted by inquiring about the ontology of goods pursued in feasts
and friendship. The last chapters offer a sketch of possible ways to re-
frame certain collective dilemmas by stressing, on the one hand, the piv-
otal role of relation goods such as fairness and trust, and on the other,
the necessity of overcoming the individual empowerment paradigm re-
garding disinformation. Finally, it shows how promoting we-reasoning
skills through the humanities (Nussbaum 2010) is not necessarily at odds
with an education focused on economic and technical skills.

Let me conclude by suggesting some contributions of this thesis to
different areas of the ontology of economics, collective action theory, and
collective responsibility. By showing a viable path regarding the ontolog-
ical irreducibility of certain collective goods and their dependence on co-
operation in we-mode, I argue that the ontological robustness of private
goods is based on the existence of specific public and relational goods.
This interdependence paves the way for an interdisciplinary and organic
understanding of collective actions and duties, which would be capable
of coherently framing all the different economic and non-economic goods
at stake. It is reasonable to highlight the constitutional linkage between
private and public moral spaces (suggesting the de-individualization of
collective dilemmas) and how education seems to be a precondition for
coherent and intertwined we-reasoning practices.

REFERENCES
Anomaly, Jonny. 2018. “Public Goods and Education.” In Philosophy and Public Policy,
Garnett, Robert F., Lenore T. Ealy, and Paul Lewis, eds. 2015. Commerce and Community:
Hale, Thomas, David Held, and Kevin Young. (2013) 2014. Gridlock: Why Global Cooper-
ation Is Failing When We Need It Most. Reprinted. Cambridge and Malden, MA: Polity
Press.


**Marco Emilio** is an Assistant Professor of Philosophy at IUSVE/Salesian Pontifical University of Venice. His current research interests lie between theoretical inquiry in ontology and emerging issues in applied social sciences. Specifically, he is mainly interested in analyzing the status of social goods and collective agency in cooperative and competitive processes.

Contact e-mail: <m.emilio@iusve.it>
PHD THESIS SUMMARY:
Defining Exploitation

ULYSSE LOJKINE
PhD in Philosophy, October 2023
Université Paris-Nanterre

While we are witnessing a resurgence in criticism of capitalist exploitation in the political sphere today, philosophers and economists approach the topic rarely and not without suspicion. For a long time, this research was dominated by the classical Marxist conception of exploitation. That is, the employer exploits the wage-earner by appropriating a share of the labor she provides. But this conception gradually receded due to an overall change in the ideological atmosphere, but also because of the technical debates of the 1970s and 1980s, led in particular by the analytical Marxism school, which cast doubt on the very possibility of measuring the flow of surplus labor. ‘Dialectical’ and ‘structuralist’ Marxists always rejected these criticisms, but failed to propose an operational definition that could lead to concrete analyses. The moral philosophy of exploitation came to the fore in the same period, but often confined itself to the study of individual choices, foregoing analysis of social structures. The aim of this thesis is to defend the possibility of a rigorous discourse on capitalist exploitation applicable to the contemporary economy. This project is inspired by both analytical and continental Marxism and seeks to reconcile them, without shying away from amending Marx's theory where this is fruitful. I first explore the positive (part I), then the normative aspects (part II) of these debates.

In part I of this thesis, chapter 1 first recalls the debates on the classical Marxist approach of measuring exploitation as surplus labor, based on the gap between the flows of labor contributed and labor commanded. Measurement of surplus labor in that sense is possible (Okishio 1963), but is overly sensitive to the choice of one among several equally plausible established conventions for the comparison of heterogeneous types of labor: if every hour of work is counted the same, well-paid workers (either in a national or in an international sense) will be exploiters, whereas if the wage is accepted as a proxy for labor complexity, wage income can never be categorized as exploitative. One way to interpret this
indeterminacy is that there is an irreducibly normative element at work in these conventions, implicitly determining how much inequality is admissible in a non-exploitative society. This leads me to follow Roemer (1982)'s counterfactual approach: exploitation analysis should be based on a socialist counterfactual; those who would gain from these counterfactual institutions compared to extant ones are exploited, those who would lose are exploiters.

The quantitative question of how to measure exploitation is followed in chapter 2 by the qualitative search for the institutions in which relations of exploitation exist. The key Marxist claim in this regard is that exploitation happens primarily in the sphere of production and not in the sphere of circulation, i.e., through wage rather than other market relations. I dispute that claim theoretically and empirically, insisting on the historical and contemporary importance, under capitalism, of two phenomena: on the one hand, direct exploitation of households through non-wage market transactions, such as household credit and home rental; on the other hand, indirect chains of exploitation mediated by multiple relations, especially when a worker is wage-exploited by a small employer who is in turn subordinated to a more powerful exploitative agent through commercial, rentier or financial relations (Lichtenstein 2012). This leads me to reject the idea that the understanding of exploitation should be grounded in the analysis of economic relations and modes of production. I argue that the latter concepts should be replaced with relations and modes of coordination, a mode of coordination being characterized by distinct relations of coordination at various scales. In the case of capitalism, this entails hierarchical coordination inside the firm and market coordination between firms, but it can also include many intermediate forms such as incentive-based management within the firm and asymmetric supply networks between firms.

In chapter 3, emphasis on the importance of large-scale indirect market interdependency allows me to shed some light on another topic: the relation between exploitation and power. In classical Marxist theory, capitalist exploitation is related to power on two levels: its direct means are ‘despotism’ inside the firm, and its structural condition of possibility is the ‘mute compulsion’ to work generated by the wage dependency of the proletarians (Mau 2023). But there is a competing neoclassical account, based on the intuition of competitive discipline. That is to say, on the labor market as in any other market, competition limits arbitrary power,
so that in the limit case of perfect competition, power relations are eliminated. I argue that a step toward solving the contradiction between the two approaches can be made by formal power analysis. I take inspiration from an idea in voting theory—that of power as decisiveness—and generalize it to contexts with strategic interactions, so that an agent has power over an outcome if it is sensitive to variations in the agent’s preferences, allowing for the other agents to strategically adapt to that variation (Lojkine 2022). This allows me to uncover monopoly power as only one component of economic power in markets, while also tracking the unequal distribution of power in competitive markets that stems from the unequal distribution of wealth.

These conceptual shifts have normative consequences (part II). Indeed, as I recall in chapter 4, the primacy of production over circulation was key in Marx and Engels’ theory of communism: because the endogenous transformation of capitalist production would give rise to communist social relations conceived as direct, immediate, and transparent power of society over itself and over things, there was no need for evaluating alternative institutions according to explicitly normative principles. But once this primacy is abandoned (as advocated in part I), it becomes clear that the abolition of exploitation cannot be achieved by a mere removal of capitalist relations. It can only result from a reconfiguration of the mode of coordination, and this reconfiguration must be the object of a normative debate, in accordance with the counterfactual reformulation of exploitation adopted in chapter 1.

If history cannot settle normative questions, can the experience of the exploited? This has often been assumed by some Marxists. Chapter 5 takes on that idea by emphasizing that the raw experience of the exploited is restricted to the local scale, and hence is unable to apprehend the exploitation relation as a whole. This is especially the case when that relation spreads across multiple levels as in the examples mentioned above. I illustrate this point by a study of the French ‘Socialisme ou barbarie’ group, which aimed to base both its exploitation analysis and its political horizon on ‘proletarian experience’ (Lefort [1952] 2013), but ended either reproducing from within a privileged role for the intellectual in charge of economic and political questions beyond the scale of the factory, or slipping into a purely immanent, non-political description of workers’ lives (Lojkine 2023). Hence, as much as the critique of exploitation has to be anchored in experience, it is not reducible to it. This raises
the question of the need for other kinds of coordinating institutions, beyond just economic ones. They should be political institutions that have the task of elaborating on the scattered experiences and aspirations of the exploited into a common normative horizon.

The thesis outlines and defends this updated and amended version of the Marxist theory of capitalist exploitation. If we accept it, then two questions should take center stage in Marxist scholarship: on the one hand, the historical and theoretical study of the various economic coordination relations and their properties in terms of exploitation and domination; on the other, the study of progressive political institutions as coordinating institutions for the aspirations of the oppressed.

REFERENCES

Ulysse Lojkine obtained his PhD in philosophy from the Université Paris-Nanterre in October 2023, where he is currently a Lecturer (Ater) in economics. He conducts research about exploitation and power, in theoretical and empirical economics, in the history of economic thought, and in philosophy.
Contact e-mail: <ulysse.lojkine@gmail.com>
PHD THESIS SUMMARY:
Sharing in Common:
A Republican Defence of Group Ownership

YARA AL SALMAN
PhD in Philosophy, June 2022
Utrecht University

Everywhere in the world, people make use of group ownership institutions. These institutions allow them to share forests, windmills, water basins, stores, meadows, and countless other resources they rely on to meet their daily needs. You would not know that from reading the philosophical literature on property, however. There, the focus is almost invariably on defining, defending, or criticizing individual ownership (see on this Olsen 2019). This leaves undertheorized three important questions: what is group ownership? Is it a valuable institution? And can it be preferable to other forms of ownership? These are the questions I have addressed in my dissertation.

I argue that group ownership is the position of authority, held by a group, to decide how an object may be used, within limits set by law (chapter 2). This authority extends both internally, over the members of the group, and externally, over outsiders. In the particular conception of group ownership that I defend, the relevant group is private and its authority is democratically organized. An example would be of farmers sharing an irrigation system to water their land. Their rights to use this irrigation system are then designed, authorized, and subject to change by their democratic decisions. This is different from various legal forms of co-ownership where the co-owners' individual rights are pre-determined by law. Such rights are ultimately reducible to individual property, while my conception of group ownership is not.

This institution is valuable when and because it empowers people. Or, to use the technical language I develop in my dissertation: group ownership is valuable when it realizes people’s basic non-domination (chapter 4). Domination, I argue, consists of the structurally enabled and structurally unequal ability of an agent to shape another agent’s option set on an arbitrary basis, meaning in ways that the subjected agent cannot control (Pettit 1997; Gädeke 2020). The opposite of domination is people’s secure
enjoyment of equal control over the relationships they are in, or non-domination. You enjoy the more limited ideal of basic non-domination when, firstly, you have basic capabilities, meaning capabilities you minimally require to withstand arbitrary power. Without, for instance, the capabilities to secure adequate nourishment, shelter, and reliable information, you can become problematically dependent on another agent, or vulnerable to their manipulation. Secondly, basic non-domination requires that people are in control of decisions that affect their basic capabilities, again to avoid making them problematically dependent on another agent.

From this ideal of basic non-domination, I derive two criteria that ownership institutions must meet according to republicanism (chapter 5). The basic capability criterion states that ownership institutions must help people to use resources to attain their basic capabilities. The control criterion holds that ownership institutions must place the people who rely on resources for their basic capabilities, in control of how these resources are used. These criteria constitute a comparative framework for the normative analysis of different ownership institutions. Group ownership is justified over other ownership institutions when, for instance, a shared irrigation system allows more farmers to access enough water to live off their land (thus securing various basic capabilities), while also ensuring that all farmers are equally in control over decisions that affect their access to water.

By analyzing common property regimes (CPRs) in natural and agricultural resources, I show the conditions under which group ownership institutions can meet both criteria. CPRs meet the basic capability criterion because they enable people to use resources to secure their livelihood (chapter 6). This may be surprising, in light of the influential myth that sharing is unproductive or leads to a ‘tragedy of the commons’ (Hardin 1968; see also Demsetz 1967). Empirical evidence shows, however, that people can use shared natural resources in a productive way if they can collectively set the rules on how these resources may be used themselves (Ostrom 1990; Cox, Arnold, and Tomás 2010). This is exactly what happens under group ownership. In my dissertation, I clarify the conditions under which CPRs can help more persons attain their basic capabilities than the alternatives of individual and public ownership. I furthermore argue that arguments against sharing are often based on speculation and have historically even been grounded in the self-interest of people who stood to gain from a privileging of individual over group ownership (see on this Ciriacy-Wantrup and Bishop 1975).
To meet the control criterion, CPRs must be internally democratic and regulated externally by a series of nested democratic communities, all of which ensure that people are in control of the decisions that affect their basic capabilities (chapter 7). Moreover, rules about inclusion and exclusion from group membership may not be determined solely by groups themselves, and separately, but must be subject to general rules which are determined by democratic deliberation and decision in a wider community. Nor may such rules track discriminatory norms, as this would reproduce rather than counter domination.

While the main focus is on CPRs in natural and agricultural resources, I show how the lessons learned can be extended to other organizations based on group ownership, including knowledge commons, energy cooperatives, insurance mutuals, and worker cooperatives. In all cases, theoretical speculation predicts that people cannot cooperate to meet their needs, while empirical evidence shows the contrary to be the case (Madison, Frischmann, and Strandburg 2010; Vriens and De Moor 2020; Dow 2003). It seems that democratic decision-making is instrumental for this achievement, while also being of intrinsic importance for people’s empowerment. Interestingly, group ownership structures have allowed these citizens to directly act to secure their basic non-domination, rather than being consigned to petitioning their government to do this for them.

My dissertation aims to contribute to political philosophy in three main ways. First, I hope that my conceptualization of group ownership will spur more research from many normative perspectives on what makes this institution valuable. In addition, my taxonomy of justifications of property (chapter 3), and my systematic republican framework for the comparative evaluation of ownership institutions (chapter 5) can facilitate research on all types of ownership. The most important contribution of my dissertation, however, is the normative defense of group ownership. Everywhere in the world, people use group ownership institutions to meet their needs in a democratic way. Thus, they take control of their lives together and as equals. It is time that their sharing practices gain the recognition they deserve, and that their value is defended.

REFERENCES


**Yara Al Salman** is an assistant professor in ethics and political philosophy at Utrecht University, where she also obtained her PhD in 2022 (*cum laude*). Her main research interests are in political philosophy, specifically on property, commons, freedom and domination, and democracy.

Contact e-mail: <y.alsalman@uu.nl>