

# Fairness Versus Welfare: The Limits of Kaplow and Shavell's Pareto Argument

Christopher P. Taggart

Follow this and additional works at: <http://scholarship.law.marquette.edu/mulr>

---

### Repository Citation

Christopher P. Taggart, *Fairness Versus Welfare: The Limits of Kaplow and Shavell's Pareto Argument*, 99 Marq. L. Rev. 661 (2016).  
Available at: <http://scholarship.law.marquette.edu/mulr/vol99/iss3/6>

This Article is brought to you for free and open access by the Journals at Marquette Law Scholarly Commons. It has been accepted for inclusion in Marquette Law Review by an authorized administrator of Marquette Law Scholarly Commons. For more information, please contact [megan.obrien@marquette.edu](mailto:megan.obrien@marquette.edu).

# FAIRNESS VERSUS WELFARE: THE LIMITS OF KAPLOW AND SHAVELL'S PARETO ARGUMENT

CHRISTOPHER P. TAGGART\*

*In a series of articles and a book, Louis Kaplow and Steven Shavell (KS) articulated and defended the normative approach of standard law-and-economics. KS also argued that legal analysts should think in welfare-economic terms exclusively when advising on normative social issues of tremendous import. This thesis generated controversy within the legal academic community because it implied that numerous analysts were not doing an important part of their jobs the way that they should be doing it. One of KS's main arguments featured a very plausible version of the Pareto principle. KS claimed that their Pareto argument demonstrated that any method of policy evaluation that gives any weight to principles independently of their effect on how well-off individuals become sometimes commits the evaluator to making everyone worse off. This Article argues that KS misstated what their Pareto argument demonstrated. It also argues that KS's Pareto argument provides no independent reason to endorse any part of welfare economics and thus no independent reason to adhere exclusively to welfare-economic thinking. Additionally, the Article clarifies much of what is at stake in deciding whether to adopt an exclusively welfare-economic approach to normative legal scholarship. Finally, the Article suggests that KS's central thesis is incorrect—there is an important place at the table for forms of normative analysis that diverge from a purely welfare-economic approach.*

I.	INTRODUCTION.....	662
II.	KAPLOW AND SHAVELL'S PROJECT.....	665
	A. <i>Critical-Level Thinking</i> .....	666
	B. <i>The Three Elements of Welfare Economics</i> .....	670
	1. Consequentialism.....	670
	2. Welfarism.....	681

---

\* Lecturer on Law, Harvard Law School. I wish to thank Duncan Kennedy and Steven Shavell for very helpful comments on earlier drafts.

3.	Utility as Preference-Satisfaction.....	694
4.	Recapitulating KS's Project.....	702
III.	KAPLOW AND SHAVELL'S PARETO ARGUMENT.....	703
A.	<i>Preliminaries—Pareto Efficiency and the Pareto Principle</i> ...	703
B.	<i>The Pareto Argument</i> .....	707
IV.	KS'S PARETO ARGUMENT DOES NOT FURTHER THEIR PROJECT.....	710
A.	<i>KS's Pareto Argument Provides No Reason to Believe (1)</i> ...	712
B.	<i>KS's Pareto Argument Provides No Reason to Believe (2)</i> ...	717
C.	<i>KS's Pareto Argument Provides No Reason to Believe (3)</i> ...	720
D.	<i>KS's Pareto Argument Provides No Reason to Believe (4)</i> ...	720

## I. INTRODUCTION

The impact of law-and-economics is difficult to overstate.<sup>1</sup> Beginning in the late 1990s, in a series of articles and a book, Louis Kaplow and Steven Shavell (KS) undertook an ambitious project in furtherance of the central moral/ethical<sup>2</sup> theory underlying normative law-and-economics.<sup>3</sup> More specifically, KS advocated an *exclusively*

---

1. Writing in 1993 and commenting on the American legal academy, former Yale Law School Dean Anthony Kronman observed:

In the years since 1965 no other approach to the study of law has had a comparable effect on the way that academic lawyers write and teach. Law and economics is . . . a permanent, institutionalized feature of American legal education. . . . [T]he movement's influence . . . is nearly unrivaled in some fields (corporations and commercial law) and dominant in others (torts, contracts, and property). The law-and-economics movement has transformed the way that teachers in these fields think about their subject and present it to their students. And in almost every area of law a working knowledge of economics is now required to keep abreast of scholarly developments, whether one is sympathetic to the movement or not.

ANTHONY T. KRONMAN, *THE LOST LAWYER* 166 (1993).

2. Sometimes the concepts expressed by terms such as "moral" and "morality" are distinguished from the concepts expressed by terms such as "ethical" and "ethics." This Article does not draw such distinctions. I treat the concept of moral (or ethical) goodness (or value) as primitive. Finally, the term "welfare economics" will refer to the moral/ethical theory that underlies (most) normative law-and-economics scholarship.

3. See generally LOUIS KAPLOW & STEVEN SHAVELL, *FAIRNESS VERSUS WELFARE* 79 (2002) [hereinafter KAPLOW & SHAVELL, *FAIRNESS VERSUS WELFARE*]; Louis Kaplow & Steven Shavell, *Fairness Versus Welfare: Notes on the Pareto Principle, Preferences, and Distributive Justice*, 32 J. LEGAL STUD. 331 (2003) [hereinafter Kaplow & Shavell, *Notes on the Pareto Principle*]; Louis Kaplow & Steven Shavell, *Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle*, 109 J. POL. ECON. 281 (2001) [hereinafter Kaplow & Shavell, *Any Non-Welfarist Method*]; Louis Kaplow & Steven Shavell, *Fairness Versus*

welfare-economic approach under which, for example, the analyst might take the position that retributive considerations should be of no direct relevance whatsoever when assessing what sanctions for criminal offenses should be imposed and that only deterrence and other forward-looking effects of criminal sanctions should even be considered. Alternatively, the analyst might consider notions of corrective justice irrelevant when considering whether tort plaintiffs should be compensated by defendants.<sup>4</sup> The analyst might also think that when someone makes a promise the fact that a *promise* was made (as opposed to, say, a mere statement of future intention) is not of any *intrinsic* relevance in deciding whether that promise should be enforced.<sup>5</sup> KS's main thesis boldly enjoined any legal analyst from ever deviating at all from a purely welfare-economic approach when generating and defending normative "law-and-policy" scholarship.<sup>6</sup>

Because KS were (and are) major proponents of the law-and-economics movement and because their project was so audacious, there was significant critical response.<sup>7</sup> As one writer claimed:

For several reasons, [KS's project] deserves close consideration. The first involves their stature as scholars. Kaplow and Shavell are prominent figures in the field of law and economics. . . . [T]hey have forayed deeply into political theory and moral philosophy, as well as a number of areas of

---

*Welfare*, 114 HARV. L. REV. 961 (2001) [hereinafter Kaplow & Shavell, *Fairness Versus Welfare*]; Louis Kaplow & Steven Shavell, *The Conflict Between Notions of Fairness and the Pareto Principle*, 1 AM. L. & ECON. REV. 63 (1999) [hereinafter Kaplow & Shavell, *The Conflict Between Notions of Fairness*].

4. For example, a welfare-economic analyst would not place *any* direct weight on the principle of corrective justice, according to which "individuals who are responsible for the wrongful losses of others have a duty to repair the losses." JULES COLEMAN, *THE PRACTICE OF PRINCIPLE* 15 (2001) (emphasis omitted).

5. For example, when determining whether a promise should be enforced under the law of contract, a welfare-economic analyst would not place *any* direct weight on the following:

There exists a convention that defines the practice of promising and its entailments. This convention provides a way that a person may create expectations in others. By virtue of the basic Kantian principles of trust and respect, it is wrong to invoke that convention in order to make a promise, and then break it.

CHARLES FRIED, *CONTRACT AS PROMISE* 17 (1981).

6. KAPLOW & SHAVELL, *FAIRNESS VERSUS WELFARE*, *supra* note 3, at 3–4.

7. See, e.g., Howard F. Chang, *A Liberal Theory of Social Welfare: Fairness, Utility, and the Pareto Principle*, 110 YALE L.J. 173, 177 (2000); Richard Craswell, *Kaplow and Shavell on the Substance of Fairness*, 32 J. LEGAL STUD. 245, 255–56 n.13 & 259 n.18 (2003).

substantive law. In light of their stature, their views have a serious claim to attention. A second reason lies in the barbed character of their arguments, many of them aimed “to convince legal policy analysts” to alter their research agendas. . . . Kaplow and Shavell argue that all law professors . . . will be wasting their time at best, and rendering pernicious advice at worst, until they embrace Kaplow and Shavell’s preferred version of welfare economics.<sup>8</sup>

One of KS’s most important<sup>9</sup> arguments, a formal argument that I shall call KS’s “Pareto argument,” is worthy of renewed attention. This Article (1) reconstructs KS’s project and, at some points, criticizes parts of it; (2) examines what KS’s Pareto argument shows; and (3) questions the extent to which that argument provides *any* reason to endorse *any* part of KS’s project that does not beg the question entirely.<sup>10</sup>

Part II will illuminate KS’s project by identifying the three main logically independent elements of welfare economics and connecting those elements to KS’s main thesis, which demands that legal analysts *exclusively* embrace welfare economics (as elaborated by KS). It is important explicitly to identify the logically independent parts of welfare economics because only then can one understand what welfare economics rules out and why. Part II will conclude that welfare economics cannot accommodate the common notion of legal rights as “trumps” but that it can, indirectly, accommodate concerns with distributive justice. However, the way that welfare economics (as KS elaborate it) could address concerns with distributive justice renders KS’s main thesis extremely implausible. Part III will reconstruct KS’s Pareto argument. Part IV will critically analyze KS’s Pareto argument and conclude that it in no way independently furthers their project.

---

8. Richard H. Fallon, Jr., *Should We All Be Welfare Economists?*, 101 MICH. L. REV. 979, 981 (2003).

9. Although this Article argues that—as far as KS’s project is concerned—the Pareto argument does not independently bolster their position at all, the Pareto argument is nonetheless important. There is a significant divide among scholars in the American legal academy between law-and-economics specialists and those who pursue other methodologies. The Pareto argument connects that divide to the foundations of social choice theory. Elaborating that connection exceeds this Article’s scope.

10. As I shall elaborate, KS recognize that the Pareto argument does not provide (and was not intended to provide) any independent reason to believe a number of their project’s claims. But I shall call into question (a) whether the Pareto argument shows what KS claim it does and (b) whether the argument provides any independent support to any part of KS’s project whatsoever. My conclusion will be that it does neither.

## II. KAPLOW AND SHAVELL'S PROJECT

As mentioned previously, KS's project is to champion a thoroughgoing welfare-economic approach when analyzing legal and policy issues.<sup>11</sup> If the analyst's objective is to recommend a course of action regarding a legal or policy issue, then the analyst *should*, according to KS, adhere strictly to an uncompromising welfare-economic approach.<sup>12</sup> KS use the term "Fairness" to refer to any moral theory or form of normative analysis that differs in any respect from welfare economics or welfare-economic analysis.<sup>13</sup> For example, a Fairness feature of a situation is a feature to which welfare economics would deny moral relevance. Along similar lines, Fairness reasoning or analysis is reasoning or analysis that counts as morally relevant something to which welfare economics imputes no moral relevance at all.<sup>14</sup> Thus, an alternative way to express KS's main normative contention is that legal academics and analysts should avoid Fairness thinking entirely when doing their normative law-and-policy work.

KS's project can be broken into two major (and related) elements. The first is to articulate and defend welfare economics, the normative framework of the law-and-economics movement.<sup>15</sup> Welfare economics can be structured as the combination of three independent<sup>16</sup> parts: (i) welfare-economic consequentialism,<sup>17</sup> (ii) welfarism, and (iii) a preference-satisfaction view of individual well-being (or individual

11. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 3–4.

12. *Id.* at 11.

13. *Id.* at 39–40.

14. *See id.* at 39 ("Notions of fairness have the property that evaluations relying on them are not based exclusively—and sometimes are not dependent at all—on how legal policies affect individuals' well-being.").

15. *Id.* at 5.

16. The independence here is logical. No component of welfare economics is logically entailed by any of the others. Even though (i), (ii), and (iii) are consistent with each other, one could, for example, accept (ii) and reject (i), or accept (iii) and reject (ii), or accept (i) and reject (ii), and so forth.

17. I call (i) "welfare-economic consequentialism" because, independently of any considerations concerning what makes one situation better or worse than another, there are many different forms of consequentialism (e.g., direct versus indirect, maximizing versus satisficing, etc.). I understand the form of consequentialism that is part of welfare economics to be of a particular kind. For example, welfare-economic consequentialism is direct and maximizing. Usually I shall just use the term "consequentialism." For a discussion of several different forms of consequentialism, see Walter Sinnott-Armstrong, *Consequentialism*, STAN. ENCYCLOPEDIA PHIL., <http://plato.stanford.edu/entries/consequentialism/> [https://perma.cc/94DD-4TU4] (last modified Oct. 22, 2015).

utility).<sup>18</sup> The second element of KS's project is the defense of a bold normative claim directed toward legal academics and similar analysts.<sup>19</sup> KS's "central claim is that the welfare-based normative approach should be exclusively employed in evaluating legal rules. That is, legal rules should be selected entirely with respect to their effects on the well-being of individuals in society."<sup>20</sup> Put a little differently, KS argue that law-and-policy academics and analysts should think on a welfare-economic, "critical" level exclusively when engaged in normative scholarship concerning social choices; they are to avoid entirely other forms of thinking, reason-giving, and arguing when generating and justifying such scholarship. That such an audacious normative thesis might be piquing (and of special interest to the legal academy) is unsurprising. If true, it would immediately follow that many legal academics have not been doing (and are continuing not to do) an extremely important part of their job the way that they should.<sup>21</sup>

#### A. Critical-Level Thinking

Before elaborating the three elements of welfare economics and discussing what each allows as acceptable forms of normative legal and policy analysis, I shall explicate "critical-level" thinking since the claim that analysts should engage exclusively in critical-level, welfare-

---

18. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 18–24. As this Article uses the term, welfare economics refers to a moral theory that provides at least partial answers to questions such as the following: What must be true for a policy choice to be ethically permissible? What makes one situation ethically better than another? What must be true of someone for her to be better or worse off than she currently is?

19. KS also consider moral reasoning by "laypersons" and official government decision makers. *Id.* at 64. But for the purposes of this Article, only KS's views about what legal academics and analysts should do are of central concern:

[W]e wish to emphasize that nothing in our argument suggests that social norms—many corresponding to notions of fairness—are inappropriate in regulating everyday life. . . .

Because we embrace welfare economics and reject giving weight to notions of fairness only when they are taken as independent principles to be used in assessing legal policy, it should be clear that our critique . . . is directed toward legal academics and other policy analysts . . . .

*Id.* at 79.

20. *Id.* at 3–4.

21. KS leave no doubt about the enjoining nature of their position: "Our thesis is entirely normative in nature; it is a claim about how legal policy analysis should be performed, not a (positive) claim about how such analysis actually is undertaken or about the content of existing legal doctrine." *Id.* at 4 n.3.

economic thinking is KS's central, distinctive thesis. KS's main thesis implicitly presupposes two different types, or levels, of moral thinking—intuitive and critical.<sup>22</sup> Intuitive-level thinking is non-inferential and often emotionally infused. For example, when presented one variant of the well-known “Trolley Problem,”<sup>23</sup> many intuitively form the moral judgment that it would be impermissible to push a person off a bridge to his death to save five others from certain doom; the calculation that five lives outweigh one is overridden by a strong sense that it is just plain wrong to kill another by shoving him off a bridge.<sup>24</sup> In contrast, critical-level moral thinking is explicitly consequentialist.<sup>25</sup> The critical-level thinker calculates—she identifies every feasible option, determines the consequences of each, and then bases her decision solely on which consequences are the best, for only that decision is morally permissible.<sup>26</sup> And not just any type of critical-level thinking is acceptable for KS. Only such thinking that tracks welfare-economic criteria for what makes a situation morally valuable is allowed.<sup>27</sup>

---

22. For a detailed explanation of the two levels, see R.M. HARE, *MORAL THINKING: ITS LEVELS, METHOD, AND POINT* (1981).

23. For a discussion of the Trolley Problem, see Judith Jarvis Thompson, *Killing, Letting Die, and the Trolley Problem*, 59 *MONIST* 204 (1976).

24. *Id.* at 206.

25. See HARE, *supra* note 22, at 44–45. There is no such thing as non-consequentialist critical-level thinking. But corresponding to each distinctive consequentialist Fairness theory, there is a distinctive form of Fairness-based critical-level thinking.

26. See generally Amartya Sen & Bernard Williams, *Introduction to UTILITARIANISM AND BEYOND* 1, 15 (Amartya Sen & Bernard Williams eds., 1982).

27. To illustrate that there are, in principle, many different types of critical-level thinking, consider the bizarre example of “welfare-taterist” critical-level thinking, which proceeds by the lights of a preposterous moral theory, which we can call “welfare-taterism.” The welfare-taterist is a consequentialist and a welfarist who holds a weird (and obviously erroneous) view of individual utility according to which any individual's well-being is an increasing function only of the number of potatoes that are located within the Grand Canyon. According to the welfare-taterist, one way to be sure to increase social welfare is to increase the number of potatoes in the Grand Canyon. (Indeed, the welfare-taterist sincerely (and consistently) thinks that every time we increase the number of potatoes that wind up in the Grand Canyon, we accomplish a Pareto improvement since everyone becomes better off.) A welfare-taterist critical-level thinker would identify society's options; determine what the outcomes of those options would be; calculate the number of potatoes that wind up in the Grand Canyon for each outcome; and then conclude that society should take the option that maximizes social welfare (by maximizing the number of potatoes that wind up in the Grand Canyon). The welfare-taterist would likely offer specific law-and-policy suggestions. Depending on the details of the analysis, such suggestions might include subsidizing potato farming; raising taxes to fund more highways leading to the Grand Canyon; funding research into genetically engineering potatoes that can be crammed more efficiently into the Grand Canyon's nooks and crannies; adding protections against legal liability for potato haulers to

Crucial to the defense of KS's main thesis is a line of reasoning developed by R.M. Hare (among others).<sup>28</sup> KS recognize that "it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim."<sup>29</sup> Put another way, welfare economics could be *indirectly self-defeating*.<sup>30</sup> A moral theory, T, is "indirectly self-defeating when it is true that, if we try to achieve our T-given aims, these aims will be worse achieved."<sup>31</sup>

For a consequentialist, the best moral decision procedure for an agent in a set of circumstances is the one that, if adopted by that agent in those circumstances, would lead to choices that have the best results. Deciding which decision procedure to follow is itself an extremely important choice. And like all choices, for a consequentialist what matters is only how good (comparatively) the consequences of the competing options are. But the decision procedure with the best consequences might require not consciously aiming at the best results.<sup>32</sup> Because we would like a defensible moral decision procedure for *us*, we need to know how to tell when to engage in critical-level thinking and when to engage in intuitive-level thinking. How do we figure that out? Referring to an ideal critical-level thinker (not subject to human limitations) as the "archangel" and an intuitive-level thinker (susceptible to human weaknesses to an extreme degree) as the "prole," Hare puts it this way:

'When ought we to think like archangels and when like proles?'  
Once we have posed the question in this way, the answer is obvious: it depends on how much each one of us, on some particular occasion or in general, resembles one or the other of these two characters. There is no philosophical answer to the

---

make trucking spuds more attractive; and so forth. The welfare-taterist would be violating KS's normative injunction even though she would be engaged in critical-level thinking. For she would not be engaged in *welfare-economic* critical-level thinking.

28. See HARE, *supra* note 22.

29. HENRY SIDGWICK, *THE METHODS OF ETHICS* 413 (photo. reprint 1962) (7th ed. 1907).

30. Derek Parfit, *Is Common-Sense Morality Self-Defeating?*, 76 J. PHIL. 533, 534 (1979).

31. *Id.* (italics removed).

32. "[T]he principles which we have to follow if we are to give ourselves the best chance of acting rightly are not definitive of 'the right act'; but if we wish to act rightly we shall do well, all the same, to follow them." HARE, *supra* note 22, at 38.

question; it depends on what powers of thought and character each one of us, for the time being, thinks he possesses.<sup>33</sup>

So for archangels, the best way to reason morally is to reason directly by the lights of whichever moral theory provides the correct criteria for whether a choice is permissible, for whether one situation is ethically better than another, and so on. KS think that the moral theory that fits that bill is welfare economics. So, KS would argue that archangels should always adhere exclusively to welfare-economic, critical-level thinking. But KS argue that we flawed human beings very often should not engage in welfare-economic, critical-level thinking.<sup>34</sup> For example, KS do not generally advocate such thinking by “laypersons” as they make moral (yet quotidian) decisions throughout their lives.<sup>35</sup> If laypersons tried to obey KS’s main thesis, they would very often get things wrong, either because of their limited rationality or because they might “cook the books” in their own favor.

But KS argue that legal analysts should adhere to KS’s main thesis for the straightforward reason that legal analysts, when they are doing their normative academic work, sufficiently approximate archangelic performance levels.<sup>36</sup> In such contexts, legal analysts and academics constitute the group of individuals who most closely resemble archangels instead of proles. They are sufficiently well-trained; they have adopted a peer-review system to assure quality-control; and they have sufficient time. Because this reasoning underlies KS’s normative (methodological) injunction to legal academics, whether KS’s main thesis is defensible turns crucially on whether welfare economics is. If welfare economics provides fallacious normative criteria, then even archangels should eschew it. So if erroneous, there would be no reason to think that legal academics should embrace welfare economics exclusively, even when legal academics resemble archangels. Thus, KS’s elaboration and defense of welfare economics is crucial to their project. All three elements of welfare economics—consequentialism, welfarism, and the preference-satisfaction view of individual utility—must be defensible if KS’s main thesis is to have any chance of being defensible. The next part examines those three elements.

---

33. *Id.* at 45.

34. See KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 68.

35. *Id.*

36. *Id.* at 389.

### B. *The Three Elements of Welfare Economics*

To more fully unpack KS's main normative thesis—that legal and policy analysts are to cleave exclusively to *welfare-economic* critical-level thinking—it is necessary to consider the three tenets that welfare economics comprises in more detail. Otherwise, the implications of KS's main thesis could not be appreciated.

#### 1. Consequentialism

According to consequentialism, (a) only the consequences of implementing feasible options are ever relevant to what choice ought to be made and (b) for a feasible option to be morally permissible, the consequence that taking that option brings about must be no worse than the outcome of any other feasible option.<sup>37</sup> A choice's consequences constitute the situation that the choice brings about. Shifting to talk of "actions" (instead of "choices"), we can formulate an extremely accommodating notion of an act's consequences: "An act's outcome [can be] construed broadly so as to include everything that would be the case were the act to be performed."<sup>38</sup> I shall assume that welfare economics embraces this very broad idea of what a policy choice's consequences might include.<sup>39</sup> Further, "[b]y itself, consequentialism is not a particular moral theory, but it becomes one, or a part of one, when combined with a theory of the good, a theory about which [situations] are better or worse than which others."<sup>40</sup>

In light of this initial characterization of consequentialism, what does consequentialism, standing alone, commit welfare economics to? Put another way, what, if anything, does consequentialism itself (independently of welfarism or the preference-satisfaction view of utility) rule out as normatively irrelevant? One possibility is that consequentialism rules out the recognition of rights if rights are

37. Sinott-Armstrong, *supra* note 17.

38. Douglas W. Portmore, *Consequentializing Moral Theories*, 88 PAC. PHIL. Q. 39, 39 (2007). Along similar lines, consequentialism "can . . . be flexible about what is to count as a 'consequence.' Any considerations which surface in the difference it makes to the world that the action was done may count towards the value of the action, even if we would not normally call these 'consequences' of the action." JONATHAN DANCY, *MORAL REASONS* 167 (1993).

39. Note that a welfare economist might think that an outcome includes many things of no relevance to the outcome's ethical value. Welfarism, to be discussed later, is about which features of an outcome can have any bearing on an outcome's ethical value.

40. Campbell Brown, *Consequentialize This*, 121 ETHICS 749, 754 (2011). Welfare economics includes a partial theory of the good—welfarism—which will be elaborated *infra* Part II.B.2.

understood as

trumps over some background justification for political decisions that states a goal for the community as a whole. If someone has a right to publish pornography, this means that it is for some reason wrong for officials to act in violation of that right, even if they (correctly) believe that the community as a whole would be better off if they did.<sup>41</sup>

Consequentialism seems to rule out respecting rights for this sort of reason. Consequentialist theories (perhaps most notably utilitarian theories) purport to provide the sort of background justification to which the preceding passage alludes. Thus, an uncompromising consequentialist would not recognize rights (understood as trumps) as normatively relevant to any social choice; the consequentialist background justification would exhaust every consideration of normative relevance. Indeed, if the outcome of respecting A's right to X would be worse than the outcome of not doing so, then it would not even be permissible to respect A's right to X—we would be required to disregard A's right to X.

But might a capacious understanding of consequentialism be able to accommodate the recognition of rights (as trumps), initial appearances to the contrary? Central to consequentialism is the idea that only “consequences” that feasible choices would “bring about” are ever of moral relevance to any social choice.<sup>42</sup> Given our (assumed) broad notion of a consequence, is there a broad notion of bring about that would allow a consequentialist to count as ethically relevant “intrinsic” or “deontological” properties of a choice? The strategy here is to find a plausible way to understand a choice as somehow included within its own consequences. If we could do that, then perhaps the intrinsic or deontological properties of a choice could also be included within that choice's consequences and therefore within the sphere of moral relevance. For example, imagine that we are considering the only two currently feasible social choices: C1 and C2. C1 would respect A's right to publish pornography, and C2 would not; that is, C1 has what might be termed a deontological, “rights-respecting” feature that C2 lacks. If a choice can be included within its own consequences and if a situation's including rights-respecting choices counts as a feature that makes that

---

41. Ronald Dworkin, *Rights as Trumps*, in THEORIES OF RIGHTS 153, 153 (Jeremy Waldron ed., 1984).

42. See Sinott-Armstrong, *supra* note 17.

situation significantly more valuable,<sup>43</sup> then, on *consequentialist* grounds, C1 might be morally required (at least in part) *because* it would respect A's right to publish pornography, while C2 would not.<sup>44</sup>

Incorporating a choice into its own consequences might also be thought to allow a consequentialist to think of evaluating a policy choice as equivalent to (or at least an aspect of) evaluating the consequences that the choice brings about. (This will become very important later in evaluating KS's Pareto argument because, as will be discussed, KS claim that the Pareto argument demonstrates a conclusion about *policy choices*.)<sup>45</sup> A social choice's consequences (outcome) can be represented formally to include the choice itself.<sup>46</sup> If we let O represent the part of the outcome of choice C that does not include C itself, we can represent the "extended" consequences of C as the pair  $\langle C, O \rangle$ . Thinking of C's outcome as  $\langle C, O \rangle$  allows us to evaluate C as part of evaluating C's outcome. If we refer to "actions" instead of "choices" and assume a (partially) hedonistic view of what makes outcomes valuable, then the following captures the basic idea behind this formalism:

Many [teleological] views hold . . . that actions themselves can have intrinsic value. On such a view, if I act in a way that gives a certain person pleasure, then the state of affairs that is realized

---

43. Welfarism, to be discussed *infra* Part II.B.2, would rule out placing direct weight on whether rights were respected when evaluating an outcome. But the current issue is whether consequentialism *itself* rules out any normative relevance for rights. In other words, the current issue is whether there is a possible consequentialist theory that would care about respecting rights as trumps.

44. This idea is reminiscent of Robert Nozick's "utilitarianism of rights":

[A] theory may include in a primary way the nonviolation of rights, yet include it in the wrong place and the wrong manner. For suppose some condition about minimizing the total (weighted) amount of violations of rights is built into the desirable end state to be achieved. We then would have something like a "utilitarianism of rights"; violations of rights (to be minimized) merely would replace the total happiness as the relevant end state in the utilitarian structure.

ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 28 (1974) (emphasis omitted).

45. The following is a preview of KS's claim about what their Pareto argument shows: "We have demonstrated that any method of *policy assessment* that is not purely welfarist violates the Pareto principle. That is, *policy evaluation* that gives any weight to principles independently of their effect on individuals' utilities will sometimes lead to *choices* under which everyone is worse off." Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 284 (emphasis added).

46. I thank Steven Shavell for pointing this out in correspondence. E-mail from Steven Shavell to Christopher P. Taggart (on file with author).

consists, among other things, in the occurrence of this pleasure and the fact that it was brought about by that action. The value of that state of affairs then depends on the value of that pleasure and possibly also on the value, positive or negative, of the action.<sup>47</sup>

Along similar lines,

[i]f . . . an agent does the right thing, he does the best of the alternatives available to him . . . . Standardly, the action will be right in virtue of its causal properties, of maximally conducing to good states of affairs. Sometimes, however, the relation of the action to the good state of affairs may not be that of cause to effect—the good state of affairs may be constituted, or partly constituted, by the agent’s doing that act.<sup>48</sup>

As the immediately preceding passage suggests, there *is* a broad idea of bring about that seems to enable welfare-economic consequentialism to understand a policy choice to be included within its own outcome. There are different ways of understanding the bring-about relation between a choice and its outcome. Perhaps the most intuitive is causal: C brings about O just in case C causes O. In discussing the idea of a necessary connection between a cause and its effect David Hume states:

When we . . . consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. . . . [T]here is not, in any single, particular instance of cause and effect, any thing which can suggest the idea of power or necessary connexion.

From the first appearance of an object, we never can conjecture what effect will result from it. But were the power or energy of any cause discoverable by the mind, we could foresee the effect . . . and might, at first, pronounce with certainty concerning it, by mere dint of thought and reasoning.<sup>49</sup>

---

47. T.M. SCANLON, *WHAT WE OWE TO EACH OTHER* 80 (1998).

48. Bernard Williams, *A Critique of Utilitarianism*, in *UTILITARIANISM: FOR AND AGAINST* 77, 86–87 (J.J.C. Smart & Bernard Williams eds., 1973).

49. DAVID HUME, *An Enquiry Concerning Human Understanding*, in *AN ENQUIRY CONCERNING HUMAN UNDERSTANDING AND OTHER WRITINGS* 3, 59–60 (Stephen Buckle ed., 2007).

As Hume explains, causal relations are not, as it were, traceable as inferences by the mind.<sup>50</sup> In contrast, logical relations are traceable “by mere dint of thought and reasoning.”<sup>51</sup> To capture this difference, Hume categorizes logical or mathematical relations as “Relations of Ideas,” in contrast to causal relations, which are “Matters of Fact”:

All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, relations of ideas, and matters of fact. Of the first kind are . . . every affirmation which is either intuitively or demonstratively certain. . . . Propositions of this kind are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe. . . .

Matters of fact, which are the second objects of human reason, are not ascertained in the same manner; nor is our evidence of their truth, however great, of a like nature with the foregoing. The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality. That the sun will not rise tomorrow is no less intelligible a proposition, and implies no more contradiction than the affirmation, that it will rise.<sup>52</sup>

If we accept Hume’s sharp distinction between the causal and the logical, then the notion of bring about that might allow a choice’s consequences to include the choice itself would embrace logical, conceptual, or constitutive relations as well as (narrowly) causal ones. To illustrate, consider the following instances of one or more actions or events bringing about an outcome:

It is an outcome of X’s hitting Y that Y is hit (by X). It is a consequence of her writing ‘c’ followed by ‘a’ followed by ‘t’ that she wrote ‘cat’. As a result of my having been born on the 8th of March, I was born a Pisces. I unlocked the door because I turned the key. By giving birth, his sister brought about his becoming an uncle.<sup>53</sup>

In each of the foregoing the outcome is, in some logical or conceptual sense, entailed by or constituted by the action or event that

---

50. *Id.*

51. *Id.* at 60.

52. *Id.* at 28–29 (emphasis removed).

53. See David Sosa, *Consequences of Consequentialism*, 102 MIND 101, 101–02 (1993) (emphasis removed).

brings it about. It is logically impossible for X to hit Y without Y being hit; (given what counts as writing the word “cat”) it is impossible to write “c” followed by “a” followed by “t” without writing that word; (given the definition of “being a Pisces” and the “8th of March”) it is impossible to be born on March 8 and not be a Pisces; (when the turning of a key is or constitutes the unlocking of a door) the one cannot occur without the other because there is essentially one event being described two different ways; finally, (given what counts as being an uncle) it is impossible for a man not to become an uncle when his sister gives birth. In these cases because of a logical, conceptual, or constitutive relationship between action and outcome, the bring-about relationship is, as Hume might say, “intuitively or demonstratively certain.”<sup>54</sup> In different ways, someone with the right concepts could, “by mere dint of thought and reasoning,”<sup>55</sup> see that the outcomes in such cases are necessarily brought about.

The upshot of the discussion to this point is a form of consequentialism that (a) features very inclusive notions of outcome and bring about and (b) seems to provide a way for a choice to be included within its own outcome. As mentioned earlier, this possibility may be thought (i) to allow a consequentialist to count deontological, rights-respecting properties of choices as normatively relevant and (ii) to show how a welfare economist can consider evaluations of policy choices to be aspects of evaluations of outcomes of those policy choices.<sup>56</sup> In the remainder of this part, I shall argue that even such an accommodating form of consequentialism (1) cannot fully appreciate the normative significance of rights as trumps and (2) can reduce the assessment of policy choices to an aspect of assessing the outcomes of those choices only by making an assumption that is extremely controversial at best.

The main reason behind (1) and (2) (immediately above) is that choices are made by choosers (agents), and agents often seem to have what have been referred to as *agent-relative reasons* to make the choices they do.<sup>57</sup> Such reasons have also been termed *non-impersonal reasons*.<sup>58</sup> Choosers are situated—their decisions are always made from

---

54. HUME, *supra* note 49, at 28.

55. *Id.* at 60.

56. *See supra* pp. 670–73.

57. PAUL HURLEY, *BEYOND CONSEQUENTIALISM* 14 n.14 (2009).

58.

[S]ome practical reasons are fundamentally non-impersonal. We have reasons to

a “point of view” or “particular context,” which (in cases where the chooser is an individual agent) is usually constituted in part by plans, projects, interests, and commitments.<sup>59</sup> Further, groups, and not just individuals, can be situated choosers.<sup>60</sup> And when a social choice, such as a policy choice or a choice of legal rule is made, it is made by a particular situated chooser (which is itself a group)—viz., society.

A very rough way to express the key idea here is that a practical reason (explanatory or justificatory) for society’s making a particular choice might be “relativized” to the point of view of society when society makes that choice. And that relativization can make a significant normative difference, which consequentialism cannot accommodate.<sup>61</sup> By way of terminology, when a reason is relativized in

---

pursue courses of action that are not based on the impartial evaluation of states of affairs, reasons that are sometimes sufficient to pursue courses of action that will not bring about the best overall consequences. . . .

On virtually every account of practical reason and deliberation . . . agents have non-impersonal reasons to act, reasons that are not themselves grounded, even indirectly, in appeals to the impersonal value of states of affairs. . . . One form of this claim that some practical reasons are fundamentally non-impersonal is a claim that the plans, projects, interests, and commitments of agents have rational significance independent of whatever rational significance they have in the determination of the best overall state of affairs, independent significance that manifests itself at the most fundamental level of practical reason.

*Id.* at 13–15.

59. *Id.* at 14 n.14.

60. For a group to make a choice there must be at least one individual in that group exercising her agency. An individual paradigmatically exercises her agency as the result of some complex psychological process having cognitive, affective, and conative dimensions. A group paradigmatically makes a choice as a result of a very different kind of process (perhaps a set of complex procedures involving voting, debate, etc.) that its members engage in. The group’s process supervenes, ultimately, on the exercises of agency by the group’s members who are participating in the group’s decision-making procedures.

61. For example, Stephen Darwall has fairly recently argued that many legal and policy issues can properly be understood only by acknowledging a type of non-impersonal reason, which he refers to as a “second-personal reason”:

I argue that there is a distinctive reason for acting—a *second-personal reason*—that is conceptually implicated in many central moral notions . . . . Reasons of this kind always involve an accountability relation between addresser and addressee—that is, that the addressee is answerable to the addresser in some way . . . . Although the claims I argue for . . . concern *morality*—moral obligation, responsibility, rights, and so on—they would also seem to bear on the *law*. One way of viewing my project . . . is that it tries to bring out the distinctive character of that part of morality that is modeled on the idea of law.

Stephen Darwall, *Law and the Second-Person Standpoint*, 40 LOY. L.A. L. REV. 891, 891 (2007).

this way, I shall refer to it as a “positional reason.” I am declining to use “agent-relative reason” and “non-impersonal reason” because I do not want to imply that society literally is an agent or person in every respect. But I do claim that a society is like an agent or a person insofar as it is a chooser that *literally can* have positional reasons (at least of a justificatory kind) to make decisions in various ways.

To illustrate a societal positional reason, imagine that a government agency (the Agency) makes a promise to Fred. (This illustration assumes that the Agency is acting on behalf of society—that the Agency is society’s agent.) Assume that if the Agency keeps its promise, then, as a consequence, three other people will break their promises to others. And assume that if the Agency breaks its promise, then the other three will keep theirs. (To simplify matters, assume that the Agency knows all of this.) Many would think that the Agency has a positional, moral reason to keep its promise, and if that reason were decisive, then the Agency ought to keep its promise to Fred. But also, let us suppose, the Agency has a moral reason to maximize promise keeping. If that reason were decisive, then the Agency ought to break its promise to Fred. The possible rights-respecting form of consequentialism mentioned previously would not recognize as morally relevant the Agency’s positional reason to keep its promise in this illustration.<sup>62</sup> Standard forms of consequentialism, such as welfare economics, evaluate choices solely in terms of outcomes.<sup>63</sup> And they evaluate outcomes without regard to the choosing agent’s particular point of view or “position within” that outcome:

[V]alue is determined impersonally; the real value of any state of affairs does not depend on the point of view of the agent. Consequently, no reference to the agent or her position in the world need enter into a [welfare-economic] consequentialist understanding of what makes an action right or wrong. Features of the particular agent may be morally relevant, but only in so far as they bear either on which state of affairs will be best or on what range of actions are open to the agent.<sup>64</sup>

---

62. See *supra* pp. 670–72.

63. See Sinott-Armstrong, *supra* note 17.

64. David McNaughton & Piers Rawling, *Agent-Relativity and the Doing-Happening Distinction*, 63 PHIL. STUD. 167, 167–68 (1991).

Critical-level (i.e., consequentialist) thinking cares nothing about the point of view of the thinker.<sup>65</sup> Archangels are ideal critical-level thinkers in part because they never “occupy” their own points of view when reasoning about what social choice should be made. Indeed, archangels might best be considered “ideal observers” who lack any particular point of view; they occupy what Thomas Nagel has referred to as a “view from nowhere.”<sup>66</sup> As mentioned earlier, unlike us human beings, archangels are capable of optimizing, impersonal objectivity.<sup>67</sup> But if the ideal critical thinker fails to recognize the moral relevance of positional reasons, then it is plausible to think that critical-level thinking systematically ignores something of potentially significant moral import:

Some impersonal moral theories . . . [hold] that we should try so far as possible to transform ourselves into instruments for the pursuit of the general good, objectively conceived (though our own interests play their part along with everyone else’s in defining that good). But while transcendence of one’s own point of view in action is the most important creative force in ethics, . . . its results cannot completely subordinate the personal standpoint . . . . The good . . . includes irreducibly subjective elements.<sup>68</sup>

One way to make the idea of a distinctively positional reason more precise is to consider situations in which *indexicals* seem to make a big difference. “Indexical[s] . . . [are] linguistic expression[s] whose reference . . . shift[s] from context to context[:] [some] paradigm[] examples . . . are ‘I’, ‘here’, ‘today’, . . . ‘he’, ‘she’, and ‘that’.”<sup>69</sup> Consider the following brief episode, suggested by John Perry:

I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter,

---

65. *See id.* at 168.

66. *See generally* THOMAS NAGEL, *THE VIEW FROM NOWHERE* (1986).

67. *See supra* pp. 668–69.

68. NAGEL, *supra* note 66, at 8. As discussed earlier, KS would deny that welfare economics requires us to transform ourselves into instruments for the pursuit of the general good all of the time. Instead, KS argue that legal analysts should transform themselves into instruments for the general good when such analysts are pursuing advisory, “normative” legal and policy analysis and argument.

69. David Braun, *Indexicals*, STAN. ENCYCLOPEDIA PHIL., <http://plato.stanford.edu/entries/indexicals/> [<https://perma.cc/MQ2R-3YZ7>] (last modified Jan. 16, 2015).

the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch.<sup>70</sup>

When it dawned on Perry that *he* was spilling the sugar, he came to know something that he did not know before.<sup>71</sup> As Perry explains, this does not seem especially mysterious. “My beliefs changed, didn’t they, in that I came to have a new one, namely, *that I am making a mess?*”<sup>72</sup> But then Perry points out that, if we restrict our attention only to facts and recognize that the statements “Perry is making a mess” and “I am making a mess” express (as uttered or thought by Perry) the same true proposition we run into a problem:

When we replace [the word “I”] with other designations of me [Perry], we no longer have an explanation of my behavior and so, it seems, no longer an attribution of the same belief. It seems to be an essential indexical. But without such a replacement, all we have to identify the belief is the sentence “I am making a mess.” But that sentence by itself doesn’t seem to identify the crucial belief, for if someone else had said it, they would have expressed a different belief, a false one.<sup>73</sup>

Note Perry’s focus on explaining his own behavior. Often, reasons “are referred to in explaining, in evaluating, and in guiding people’s behaviour.”<sup>74</sup> Whatever else normative reasons do, they must be able to *explain* and *justify* the choices of moral agents. Although not designed to be specifically about moral reasons, Perry’s example shows how thinking about moral reasons exclusively in terms of their status as true propositions does not capture every respect in which moral reasons might *explain* choices. Only by considering how the choosing agent could express practical reasons to herself (with indexicals) can we obtain an *explanation* of her choice. The additional, crucial idea that the consequentialist denies is that indexicals are ever indispensable for a *justification* of a choice. To express the positional, *explanatory* reason for choice, as in the example above, we need indexicals. Similarly, to express positional, *justificatory* (i.e., at least morally relevant) reasons for choice, we need indexicals.

---

70. John Perry, *The Problem of the Essential Indexical*, 13 NOÛS 3, 3 (1979).

71. *Id.*

72. *Id.*

73. *Id.* (emphasis removed).

74. JOSEPH RAZ, PRACTICAL REASON AND NORMS 15–16 (1975).

To illustrate by returning to our previous example of rights understood as trumps: “[i]f someone has a right to publish pornography, this means that it is for some reason wrong for officials to act in violation of that right, even if they (correctly) believe that the community as a whole would be better off if they did.”<sup>75</sup> If the publisher’s right is a trump, then not only do the officials have a positional justificatory reason to allow the publisher to publish pornography but also that reason is dispositive. Further, the expression of that positional justificatory reason would require the use of indexicals. In this case, in the mouth of the officials acting in their official capacity, the reason might be expressed as, “We are forbidden to violate the publisher’s right to publish pornography.” The indexical “we” refers to the officials only insofar as they are acting on behalf of society—the “we” here effectively *is* society, and the officials’ choice not to interfere with the publisher’s activities is a social choice. Facts entailed by who has what rights are (when appropriately expressed with indexicals) positional practical reasons, which seem at least normatively relevant to (if not dispositive regarding) many social choices. Standard forms of consequentialism, such as welfare-economic consequentialism, leave no room for such reasons. So standard understandings of the normative significance of rights are ruled out by consequentialism.

Along similar lines, policy assessment can be reduced to an aspect of outcome assessment by including policy choices within their own outcomes only if there are no positional reasons of normative relevance to any policy choice. Welfare-economic consequentialism is a form of consequentialism that makes this (contentious) assumption. It seems straightforward enough to identify examples of legal or policy choices in which there would be positional reasons of normative relevance to what the government should do. Imagine that the government is deciding whether to enact a statute that would retroactively invalidate a very small number of its own contractual obligations with private parties. (Assuming that it was even legally possible, there may well be many reasons why this would be a bad idea, but these are beside the point of the example.) Imagine also that the Agency’s promise to Fred from the earlier illustration<sup>76</sup> was legally enforceable under standard contract law principles—there was mutual assent, sufficient consideration, and so forth. Additionally, imagine that the Agency’s promise is among the

---

75. Dworkin, *supra* note 41, at 153.

76. *See supra* pp. 676–77.

few promises that the statute under consideration would render unenforceable. As expressed in the “mouth” of the government the following two reasons seem to be *at least* of moral relevance to what the government should do: (1) “Parties should keep their promises”; and (2) “I made a promise to Fred.” And it follows from these two premises that (as expressed in the government’s “mouth”) “I should keep my promise to Fred.”

Thus, welfare-economic consequentialism cannot fully appreciate the normative significance of rights as trumps, and it can reduce the assessment of policy choices to an aspect of assessing the outcomes of those choices only by making the controversial assumption that there are never any positional practical reasons of normative relevance to any policy choice. Of course, one might insist that this does not conclusively demonstrate that welfare-economic consequentialism is false. Such a demonstration would require a conclusive argument that positional reasons can (at least sometimes) be of moral relevance to social choice, and I have offered only what may seem to be clear examples (such as “I should keep my promise to Fred” in the government’s “mouth”). But the foregoing does clarify some of what consequentialism, standing alone, rules out as normatively relevant to social choice. In the next part, I turn to welfarism—welfare economics’ account of what can be of any relevance to the ethical value of a situation.

## 2. Welfarism

Welfarism is welfare economics’ (partial) account of the good. As mentioned previously, a choice’s consequences constitute the state that the choice brings about.<sup>77</sup> Welfarism is a view about the value (amount of moral/ethical goodness realized by) a state.<sup>78</sup> According to welfarism, the only features of a situation that determine its value are, collectively, the situation’s *utility information*. Speaking roughly, a situation’s utility information is nothing more than information about what “bearers of utility” exist in that situation and how well-off each such bearer is in that situation. In short, for a welfarist, the ultimate source of all value realized in a situation is a function only of the well-being of individuals in that situation.

---

77. See *supra* note 39 and accompanying text.

78. See Amartya Sen, *Utilitarianism and Welfarism*, 76 J. PHIL. 463, 464 (1979) (defining welfarism as “the principle that the goodness of a state of affairs depends ultimately on the set of individual utilities in that state, and—more demandingly—can be seen as an increasing function of that set”).

Note that, as this Article uses the term, “welfarism” is not automatically a type of consequentialism. Often, welfarism is considered a species of consequentialism, such that the idea of a non-consequentialist welfarist would be a contradiction, like the idea of a non-four-sided square. But as this Article uses the term, welfarism could be combined with a non-consequentialist morality, even though it is often combined with consequentialism and welfare economics does so combine them. (Note that a non-consequentialist might think that consequences are *always* of moral relevance and that the value of any consequence (situation) is a function only of the utilities of the individuals in that situation. Of course, such a non-consequentialist would also think that there are things other than a feasible option’s consequences that *may be* relevant to whether taking that option is morally permissible.)

Once welfarism is understood not automatically as a form of consequentialism, the source of the welfare-economic idea that any situations to be considered in the course of normative analysis are the outcomes of competing feasible options can be more precisely identified. That source is consequentialism,<sup>79</sup> not welfarism. That is, there is nothing incoherent about a welfarist comparison of two entirely hypothetical (perhaps infeasible) states. Such states need not be thought of as the outcomes of any choice. Welfarism would constrain evaluative comparisons of such completely fictitious states just as it would constrain evaluative comparisons of the outcomes of doable choices.<sup>80</sup>

As recently mentioned, welfarism’s central claim is that the only features of a situation relevant to the situation’s value are, collectively, the situation’s utility information. Of course, to understand this claim,

---

79. Recall that consequentialism is the view that (a) only the consequences of implementing *feasible* options are ever relevant to what choice ought to be made, and (b) for a *feasible* option to be morally/ethically permissible, the consequence that taking that option brings about must be no worse than the outcome of any other *feasible* option. Consequentialism cares about feasibility; welfarism (at least as this Article defines it) does not. Welfare economics cares about feasibility because it is a form of consequentialism, not because it includes welfarism.

80. To clarify what I mean by the comparative value of completely fictitious states, I am not suggesting that a state of affairs has moral value as an un-actualized abstraction. When we ask about the moral value of a state of affairs that is not actual, we are asking what value it would have if it were actual. (Similarly, if I say that Linda is in a state of affairs, I mean that Linda would exist if that state were actual.) Thus, even if we were comparing situations that were not just fictional but fantastical, welfarism would constrain comparative evaluations.

one must understand what a situation's utility information is. And fully understanding what utility information is requires understanding what "individual utility" is. Part II.B.3 will consider welfare economics' view of what individual utility is. The remainder of this part will discuss a situation's utility information in abstraction from any particular view of individual utility. This will clarify what welfarism, standing alone, commits the welfare economist to, and it will help shed light on what welfarism (standing alone) might be understood to rule out as normatively relevant.

Because practitioners of normative law and economics are committed to welfarism, it is sometimes asserted that they deem concerns about distributive justice insignificant:

Perhaps the most common criticism of law and economics is that it overlooks or . . . displaces questions of distribution or equity. When analyzing the efficiency of one or another area of law, legal economists typically . . . treat distributional consequences as irrelevant. . . . Economists respond [to this criticism] in part by observing that distributional questions taken by themselves fall outside the reach of economic science.<sup>81</sup>

As this part will argue, welfarism, as articulated and defended by KS, can accommodate distributional concerns.<sup>82</sup> But as will also be argued, this accommodation comes at a price, which calls KS's main normative contention (addressed to legal academics) into serious doubt.

Conventionally, how an individual's utility is affected by the situation she is in is represented by a utility function—"a function from the set of states of the world,  $X$ , to the real line,  $R$ ."<sup>83</sup> If (of two situations A and B) Linda is better off in A than in B, then  $U_{\text{Linda}}(A)$  is a real number representing Linda's utility in A;  $U_{\text{Linda}}(B)$  is a real number representing Linda's utility in B;  $U_{\text{Linda}}(A) > U_{\text{Linda}}(B)$ ; and  $(U_{\text{Linda}}(A) - U_{\text{Linda}}(B))$  is a positive real number representing how much better off Linda is in A than in B. To speak very roughly, much as individual utility functions represent how individual utilities vary with situations, a social welfare function (SWF) represents how "social utility" varies with

---

81. Jon D. Hanson & Melissa R. Hart, *Law and Economics*, in *A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY* 311, 330 (Dennis Patterson ed., 1996).

82. See *infra* Part II.B.2; see also KAPLOW & SHAVELL, *FAIRNESS VERSUS WELFARE*, *supra* note 3, at 29.

83. Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283 (describing social welfare functions, and soon afterward, similarly characterizing individual utility functions).

situations.<sup>84</sup> Like an individual utility function, a SWF is “a function from the set of states of the world,  $X$ , to the real line,  $R$ .”<sup>85</sup> Essentially, a SWF is a situation evaluator. It takes as its input a logically possible state and yields as its output an evaluation of that state. If we let  $F$  be a SWF that places more value on state  $A$  than on state  $B$ :  $F(A)$  is a real number representing how much value  $F$  assigns to  $A$ ;  $F(B)$  is a real number representing how much value  $F$  assigns to  $B$ ;  $F(A) > F(B)$ ; and  $(F(A) - F(B))$  is a positive real number representing how much more valuable  $A$  is than  $B$ , at least according to  $F$ .

According to KS, for a SWF,  $W$ , to be condoned by a welfarist, it must have the following general form:  $W(x) = F(U_1(x), U_2(x), \dots, U_n(x))$ , where  $U_i(x)$  is the utility function for individual  $i$  from a society comprising  $n$  individuals, and  $x$  ranges over possible states.<sup>86</sup> Following common usage, KS sometimes refer to such a SWF as an “individualistic social welfare function.”<sup>87</sup> In addition to being of this form,  $W(x)$  must (1) be an increasing function of the utilities of individuals and (2) respect a plausible moral precept, which KS call the *condition of equal concern* (CEC).<sup>88</sup> Regarding (1), KS claim “[an individualistic] social welfare function can be *any* increasing function of individuals’ utilities.”<sup>89</sup> Thus, welfarism does not commit the analyst to either a standard utilitarian SWF (add up the utilities) or an “average utilitarian” SWF (add up the utilities and divide by the number of utility bearers). But, of course, welfarism *allows* utilitarian SWFs. I shall discuss (2)—CEC—in more detail shortly.

With this apparatus at hand, we can now reformulate welfarism’s central claim: A SWF of the form  $W(x) = F(U_1(x), U_2(x), \dots, U_n(x))$  evaluates a state  $x$  in light only of those features of  $x$  that are relevant to  $x$ ’s value; a SWF not of that form does not.  $W(x)$ ’s form suggests a narrow concept of a state’s utility information. For a given state  $x$ , the utility information that  $W$  has to “work with” could be represented by an ordered set of numbers, one per individual, where each number represents a different individual’s utility in  $x$ . Put another way, the only information that  $W$  countenances can be represented by a set of individual/number correlations. So if  $x$  contained exactly four

84. *See id.*

85. *Id.*

86. *See* KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24 n.15.

87. Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283.

88. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24–25.

89. *Id.* at 24 n.15 (emphasis added).

individuals<sup>90</sup>—Jill, Jack, Bill, and Bob—we might imagine  $x$ 's utility information to be represented as:  $\{\langle \text{Jill}, 5 \rangle, \langle \text{Jack}, 7 \rangle, \langle \text{Bill}, 10 \rangle, \langle \text{Bob}, 5 \rangle\}$ . Bill is best off in  $x$ ; Jill is one of the two worst off in  $x$ ; Jack is better off than Bob by “2”; and so forth.<sup>91</sup>

Turning to the condition of equal concern, CEC requires us “to go beyond ‘I’ and ‘you’ to the universal law, the universalisable judgment, the standpoint of the impartial spectator or ideal observer, or whatever we choose to call it.”<sup>92</sup> This universalizing impulse seems central to any plausible moral theory—consequentialist, deontological, or otherwise.<sup>93</sup> In Part II.B.1, this impulse was discussed in connection with consequentialism. There, the impartiality of the social *chooser* vis-à-vis the chooser's position or point of view was paramount.<sup>94</sup> In connection with welfarism, in contrast, CEC demands impartiality as to the *stakeholders*—the individual bearers of utility in the situation being evaluated.<sup>95</sup> As KS explain, “[i]t is . . . generally supposed that each individual's well-being affects social welfare in a symmetric manner, which is to say that the idea of social welfare incorporates a basic notion of equal concern for all individuals.”<sup>96</sup>

Consider one way that KS might try to accommodate concerns about distributive justice while exclusively endorsing SWFs that (i) obey CEC and (ii) are sensitive only to the utilities of individuals. To borrow an example by Steven Shavell, an acceptable welfarist SWF could amalgamate individual utilities by summing their square roots to reflect a preference that utility be distributed evenly:

[C]onsider a situation where there are two individuals, and each has the same utility, 100. Then social welfare is 20, namely,  $\sqrt{100} + \sqrt{100} = 10 + 10$ . This equal distribution of utility is superior to

---

90. That a state or situation  $x$  contains individual S means that if  $x$  were actual, then S would exist.

91. Not all utility functions and SWFs are such that their specific cardinal values matter that much. For such “ordinal” functions, assigning 10 to Bill and 5 to Jill is just a way of indicating that Bill is better off than Jill. The specific difference ( $10 - 5 = 5$ ) is not significant. For ranking purposes, assigning Bill a utility index of  $4\pi$  and Jill an index of  $3\sqrt{3}$  would work just as well. Nonetheless, this Article shall presuppose real-valued individual utility functions and SWFs whose cardinal values matter because they matter in KS's Pareto argument, which will be presented in Part III and critically analyzed in Part IV.

92. PETER SINGER, PRACTICAL ETHICS 12 (2d ed. 1993).

93. *Id.*

94. See *supra* Part II.B.1.

95. Of course, the chooser might also turn out to be a stakeholder.

96. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24–25.

the unequal distribution where one person has utility of 50 and the other of 150, in which case the social welfare is 19.32 (for  $\sqrt{50} + \sqrt{150} = 7.07 + 12.25$ ), and this distribution is superior to the extreme distribution in which one person has all the utility of 200, in which case social welfare is 14.14 (for  $\sqrt{200} = 14.14$ ).<sup>97</sup>

The SWF in this example treats the utilities of all individuals the same, and it is responsive only to the utilities of individuals. What a welfarist SWF could not do, however, is place *direct* weight on the degree to which a situation's pattern of utility distribution is egalitarian.

Things get more complicated, however. As KS elaborate their version of CEC:

With respect to [an individualistic] social welfare function . . . the condition of equal concern . . . is the stipulation that the value of W does not depend on which individual has which utility level. This assumption rules out a function that favors a specific individual at the expense of others. Formally, the equal treatment requirement is usually expressed as a symmetry or anonymity condition. For example, it is assumed that a situation in which Jill has utility of 5 and Bill has utility of 10 must be viewed as no better and no worse than one in which it is Jill who has utility of 10 and Bill utility of 5.<sup>98</sup>

Consider the following comparison of the distribution of utilities in three different states, *x*, *y*, and *z*:

	Jill's Utility	Jack's Utility	Bill's Utility	Bob's Utility
State <i>x</i>	5	7	10	5
State <i>y</i>	5	7	10	5
State <i>z</i>	10	7	5	5

KS's elaboration of CEC suggests that even though there may be numerous differences among these states, *x*'s utility information is not just the same as *y*'s; it is also the same as *z*'s.<sup>99</sup> Assuming consistency in the order of placeholders for individual utility values in the inputs of

97. STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 597 n.5 (2004).

98. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 25 n.16.

99. *Id.* at 24–25.

SWFs, KS's version of CEC *appears* to require that for any acceptable individualistic SWF  $W$ ,  $W(\langle 5, 7, 10, 5 \rangle) = W(\langle 10, 7, 5, 5 \rangle)$  because the only difference between the inputs is to "swap" Jill's utility and Bill's.<sup>100</sup> But now we are faced with a puzzle. If KS are correct that it never matters whether it is Jill or Bill who has a utility of 5 instead of 10, then certain increasing functions of individual utilities seem to get ruled out, contrary to KS's insistence that *any* increasing function of utilities could be acceptable (in principle) under welfarism. For example, the following function now seems to be ruled out as a possible individualistic SWF:  $W(\langle a, b, c, d \rangle) = a + 2b + 4c + 4d$ .  $W(\langle 5, 7, 10, 5 \rangle) = 79 \neq 64 = W(\langle 10, 7, 5, 5 \rangle)$ , in violation of KS's anonymity condition.  $W(\langle a, b, c, d \rangle)$  increases when  $a$ ,  $b$ ,  $c$ , or  $d$  do.<sup>101</sup> So  $W(\langle a, b, c, d \rangle)$  is definitely an increasing function of individual utilities. But  $W(\langle a, b, c, d \rangle)$  now appears to be ruled out by KS's version of CEC (by its anonymity condition in particular).

The key to solving this puzzle is to recognize a further refinement of KS's version of CEC: "[CEC] does not...rule out attention to differences in income, ability, opportunity, need, and the like, as all these factors affect well-being and thus the level of utility."<sup>102</sup> Thus, if, for example, Jill is less needy than Bill, then it may well be consistent with KS's version of CEC to view a situation in which Jill has utility of 5 and Bill has utility of 10 as better than one in which it is Jill who has utility of 10 and Bill utility of 5.<sup>103</sup> Furthermore, it is consistent with KS's version of CEC for a SWF to yield these differing evaluations by placing more weight on Bill's utility than on Jill's. The crucial point is that the SWF cannot place more weight on Bill's utility than on Jill's *simply because* he is Bill and she is Jill. CEC rules out such bald-faced favoritism. But if the *reason* that the SWF places more weight on Bill's utility than on Jill's is that Bill has a normatively significant property that Jill lacks (e.g., being needy), then it would be consistent with CEC for the SWF to favor Bill's utility more than Jill's. To put the same point a little differently, it is consistent with KS's version of welfarism for an analyst to make the value judgment that, say, the poor should be favored over the rich insofar as the analyst determines how much the utilities of particular individuals should count when choosing a SWF.

---

100. *Id.*

101. The "or" here is inclusive.

102. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 25 n.16.

103. *See id.* at 30–31.

KS openly acknowledge that “the approach of welfare economics involves value judgments.”<sup>104</sup> The first (unsurprising) value judgment is “that social welfare depends on individuals’ well-being, that this dependence is positive, and that factors unrelated to individuals’ well-being are irrelevant.”<sup>105</sup> The second (perhaps surprising) value judgment that welfare economics involves is necessarily reflected in the analyst’s choice of specific individualistic SWF. “The choice of a method of aggregation involves the adoption of a view concerning matters of distribution.”<sup>106</sup> (Here, KS are referring to the distribution of utility among individuals.)

This complication calls for a reassessment of what a state’s utility information is, at least according to KS’s version of welfarism. Recall that welfarism’s central claim is that the only features of a situation relevant to that situation’s value are, collectively, the situation’s utility information.<sup>107</sup> According to KS, at least some differences between people in a situation—differences *other* than their individual utilities—can be relevant to what that situation’s value is.<sup>108</sup> For example, whether some people in situation S are needier than others is a feature of S that can be relevant to a welfarist assessment of S’s value. This follows because the welfarist is allowed by KS to promote a value judgment favoring the needy in selecting which individualistic SWF to use to evaluate S. And the analyst’s selection of SWF is relevant to what S’s

---

104. *Id.* at 25.

105. *Id.* at 25–26.

106. *Id.* at 26–27. Note that KS recognize that welfare economists must make a number of different kinds of value judgments when engaged in normative legal and policy analysis:

In addition to providing a method of aggregating individuals’ utilities, a complete account of social welfare would address questions about membership in the group of individuals whose utilities are to be aggregated—whether it includes all individuals in a nation, or in the world, or in some other group; whether it includes only the present generation or also future ones; and whether it includes only humans or, for example, all sentient beings. . . . We note that these questions also must be addressed under other systems of evaluation. For example, if individual autonomy is to be honored, one must specify which individuals are entitled to such consideration. These topics . . . are beyond the scope of our inquiry, the purpose of which is to focus on the difference between evaluative principles that are denominated solely in terms of well-being and principles that are based in whole or in part on other factors—a distinction that is qualitatively similar regardless of how these other questions are answered.

*Id.* at 26 n.19.

107. See *supra* note 79 and accompanying text.

108. See KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 30–31.

value turns out to be. In short, according to KS's version of welfarism, how an individual stakeholder in a situation can accurately be described (*independently* of her utility) could affect what that situation's utility information is.

Previously, we considered an example in which a state's utility information was represented as the set  $\{\langle \text{Jill}, 5 \rangle, \langle \text{Jack}, 7 \rangle, \langle \text{Bill}, 10 \rangle, \langle \text{Bob}, 5 \rangle\}$ . But now it is clear that this representation may well have been significantly incomplete. This representation picks out individuals exclusively via rigidly designating proper names—e.g., “Jill” picks out Jill. The set is indifferent to the various descriptions that the individuals might or might not satisfy, and as just explained, some such descriptions might be relevant to the state's value. (In our previous example it was relevant that Bill was *accurately* describable as a needy person and that Jill was not.) So in at least many cases KS's version of welfarism calls for representations of utility information that might look more like the following:

$\{\langle \text{Jill (who is not needy)}, 5 \rangle, \langle \text{Jack (who is not needy)}, 7 \rangle, \langle \text{Bill (who is needy)}, 10 \rangle, \langle \text{Bob (who is not needy)}, 5 \rangle\}$

It therefore seems that, given KS's elaboration of CEC, welfarism is able to accommodate distributional concerns.<sup>109</sup> One mechanism by which the welfarist analyst can do this is by selecting which individualistic SWF to use to reflect the analyst's personal value judgments.<sup>110</sup> But although understanding welfarism this way enables KS to accommodate distributional concerns, it comes at a big price. When a welfarist analyst selects a particular individualistic SWF, he inescapably does so on the basis of a personal value judgment, which in turn is based on treating individuals as more than entirely anonymous bearers of utility.<sup>111</sup> Further, he must make *some* such selection, for otherwise no welfare-economic calculation would be possible. And if the analyst is to respect KS's main normative thesis, then he must implicitly treat individuals as more than entirely anonymous bearers of utility while engaging exclusively in welfare-economic critical-level reasoning. This means that, although welfare economists can and do make implicit (and perhaps even determinative) value judgments when

109. *Id.*

110. In the next part, this Article will consider another means by which welfare economics might try to accommodate concerns about distributive justice—recognizing “tastes for fairness.”

111. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 27–28 n.23.

selecting SWFs for normative analysis, there is one thing that they can never do *if* they utilize *only* welfare-economic critical-level thinking—they cannot even *attempt* to offer any reasoned justification whatsoever for any of their crucial value judgments.

To frame the argument here in four steps:

- (a) If doing X is impermissible for S and doing X is necessary to do Y, then doing Y is also impermissible for S. [premise]
- (b) Placing direct weight on features of individuals other than their utilities (even implicitly) when evaluating situations is impermissible for a practicing normative analyst. [premise]
- (c) Placing direct weight on features of individuals other than their utilities (even implicitly) when evaluating situations is necessary to offer a justificatory reason in support of a particular chosen individualistic SWF. [premise]
- (d) Therefore, offering a justificatory reason in support of a particular chosen individualistic SWF is impermissible for a practicing normative analyst. [from (a), (b), and (c)]

Step (d) rules out even *trying* to justify a choice of a particular individualistic SWF. The justificatory reason that (d) alludes to might be a good one, though it might not be. Step (d) validly follows from the combination of (a), (b), and (c). I am assuming that (a) is true.<sup>112</sup> Step (b) is a partial restatement of KS's main normative injunction, which (as explained) is addressed to legal academics. To implicitly place direct weight on features of individuals other than their utilities when evaluating situations that contain those individuals is to engage in forbidden Fairness thinking. Therefore, if (c) is true, then KS's position commits them to (d). So is (c) true?

Step (c) indeed is true. To see why, imagine that Jill is a "law-and-economics" professor considering what to recommend about a particular social choice of significant import. Imagine also that she is sincerely convinced (perhaps on the basis of an intuitive value judgment about entitlements) that persons of Native American ancestry should be entitled to special consideration because of past injustices perpetrated against their forebears. There are four members of society under consideration: Jill herself, Joan, Bill, and Bob. Jill and Joan are of Apache ancestry, and Bill and Bob are not descendants of Native

---

112. If this assumption is incorrect, then the argument is not sound. But (a) seems to be true.

Americans. As it turns out, there are only two feasible options available whose outcomes would be, respectively, O1 and O2. The name/number utility matrices for the two outcomes are:

	O1	O2
Jill	5	6
Joan	4	5
Bill	7	0
Bob	6	0

In light of the important social choice about which Jill is trying to give reasoned advice, Jill ponders the most defensible way to amalgamate utilities in support of whatever course of action she will ultimately endorse in a law-and-policy article that she eventually publishes. She could use an “average-utilitarian” SWF, as we could imagine her law-and-economics colleague (call him Harsanyi) urges in light of arguments that he previously published favoring that form of utilitarianism. But she finds her intuitive value judgment about the proper distribution of utilities more convincing, so in the journal article that she ultimately publishes she uses an individualistic SWF that is similar to an average-utilitarian SWF but that multiplies the utilities of descendants of Native Americans by ten before averaging to compare the two outcomes. Unsurprisingly, she evaluates O2 (to which she assigns a social welfare of 27.50)<sup>113</sup> to be better than O1 (to which she assigns a social welfare of 25.75).<sup>114</sup> As a result, in her journal article, Jill (in her capacity as policy analyst) publically recommends that society take the choice that leads to O2. If she had instead been convinced by Harsanyi, then she would have used an average-utilitarian SWF and would have evaluated O1 (which would have received a value of 5.50)<sup>115</sup>

---

113.  $((6 \times 10) + (5 \times 10) + 0 + 0) \div 4 = 27.50$

114.  $((5 \times 10) + (4 \times 10) + 7 + 6) \div 4 = 25.75$

115.  $(5 + 4 + 7 + 6) \div 4 = 5.50$

to be better than O2 (which would have received a value of 2.75).<sup>116</sup> That is, if Harsanyi had convinced her, then she would have offered the opposite advice—what she advised turned on the individualistic SWF that she chose and, ultimately, on her *reason* for that choice.

If in her article Jill offered her reason to *justify* her individualistic SWF over others (such as a straightforward average-utilitarian SWF), then she would *ipso facto* be appealing to a feature of O1 and O2 (that they contain Native American individuals whose ancestors were unfairly treated) to *justify* how she comparatively evaluates any two situations sharing that feature. Thus, she would of necessity be appealing to features of O1 and O2 other than the utilities of the individuals in those outcomes in comparing them. And that would be a paradigmatically clear instance of what KS claim is contrary to what a legal academic *should ever do*.<sup>117</sup>

Since (c) is true, KS's position commits them to (d). And (d) is an indefensible normative restriction—it is indefensible to assert that an analyst *should never even try* to justify the individualistic SWF that is selected for use in normative analysis of a serious social or legal issue. To illustrate, consider the very important social choices that the United States faces as it continues to respond to a politically-charged question: How should the U.S. structure its immigration policies? Not that long ago, economist Gary Becker published a normative recommendation in a blog saying:

The United States . . . uses quotas that give preference to family members of persons already here legally, to applicants with greater skills, to persons who applied earlier, and some other criteria. . . . [T]he best alternative to the present quota system is an ancient way of allocating a scarce and popular good; namely, by charging a price that clears the market. That is why I believe countries *should* sell the right to immigrate, especially the United States that has so many persons waiting to immigrate.<sup>118</sup>

Becker proceeded to offer a number of different consequentialist arguments in support of this idea and drew the ultimate conclusion that, assuming that the price was set in the right way, “charging a fee to

---

116.  $(6 + 5 + 0 + 0) \div 4 = 2.75$

117. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 3, 52–62.

118. Gary Becker, *Sell the Right to Immigrate*, BECKER-POSNER BLOG (Feb. 21, 2005), <http://www.becker-posner-blog.com/2005/02/sell-the-right-to-immigrate-becker.html> [<https://perma.cc/PMP8-T4JY>] (emphasis added).

immigrate would raise tax revenue, increase the number of immigrants accepted, and also raise the quality of those accepted. It is a win-win situation for countries accepting immigrants, and for the vast majority of persons who would like to immigrate.”<sup>119</sup>

The way Becker framed his conclusion might be taken to suggest (in light of the term “win-win”), not just that implementing an immigration market system would improve social welfare, but that turning to such a system would lead to a Pareto improvement over the quota system’s status quo. And Pareto improvements are hard to disagree with, especially ones that make everyone better off. This seems an implausible interpretation though. Becker was not suggesting that literally everyone would be better off. It seems very likely, for example, that as a consequence of the change in policy, a number of Americans, who would have been able to get particular high-paying jobs at home under the immigration quota system, would not get those jobs if there were an “immigration market.” Under a fee system, many better-qualified immigrants who could pay the immigration fee would get those jobs instead. Or we could imagine that at least some potential immigrants, who would have been able to come to the United States under the quota system, would not get in under the fee system because they could not muster a sufficient down-payment to entice a bank to offer them an “immigration loan” to enable them to pay the U.S. immigration market price up front.

None of this is to suggest that Becker’s proposal is a bad one. Maybe, all things considered, selling the right to immigrate is what the United States should do.<sup>120</sup> But the example brings out how the strength of Becker’s proposal implicitly depends on the choice of individualistic SWF that underwrites his analysis. It is not implausible, for example, to think that the United States might owe its own citizens a degree or type of consideration that it does not owe persons from other countries. This line of thought would not be suggesting that people from other countries hoping to immigrate do not matter at all but that the interests of current U.S. citizens matter more. And for these reasons, Becker’s evaluations could be questioned. Weight could (perhaps properly) be placed on features of outcomes other than individual utilities to back up selecting a particular individualistic SWF—for example, extra weight could be placed on the fact that some individuals affected are already Americans.

---

119. *Id.*

120. This Article remains entirely neutral on this issue.

Alternatively, extra weight might be placed on the fact that some aspiring immigrants are indigent individuals whose spouses have already lawfully been admitted to the United States. Either way, the utilities of individuals who fall under certain categories would be weighted differently than those of others, changing the individualistic SWF that is implicitly operative and trying to justify that change. Indeed, this example highlights another choice that cannot be rationalized in purely welfare-economic terms—who should count as a utility-bearer at all for the purpose of a United States national immigration policy decision? Even if an immigration market would provide a win-win outcome under which the vast majority of persons who would like to immigrate would win, for what reasons should it matter whether they wind up winning too? (This is not to suggest that it should not matter—just that this is a critical question.)

If Becker even *tried* to address these crucial questions by providing an argument or even one reason, he would be doing what KS say he *should not* do.<sup>121</sup> Any such argument or reason would, of necessity, require Becker to appeal to a rationale that was inconsistent with purely critical-level, welfare-economic thinking. I submit that this implication is indefensible. How could even *attempting* to justify such important value judgments by appeal to reason be contrary to what a normative law-and-policy analyst should do when generating normative law-and-policy scholarship? If anything, a complete failure to try to offer any such justification seems to be what the analyst *should avoid*. In sum, KS's version of welfarism can accommodate distributive concerns via the choice of SWF but only at the cost of rendering their main normative thesis extremely implausible.

### 3. Utility as Preference-Satisfaction

According to welfare economics, an individual's well-being (or utility) is that individual's degree of preference (or desire) satisfaction<sup>122</sup>: What does the individual want the world to be like, and to what extent does it seem to her that/do<sup>123</sup> things turn out that way? It bears

---

121. *See supra* Part II.A.

122. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 409.

123. The slash here tracks two different ways of understanding what satisfying a preference or desire is. For the purposes of this Article, I shall assume that an individual's desire that P is satisfied only if she believes that P and P is true. Thus, I shall assume the following: (a) a deluded individual's preferences are not satisfied when she falsely (but very happily) believes that the world is the way that she would like it to be, and (b) an ignorant

emphasis that what an individual actually<sup>124</sup> desires, not what an analyst thinks the individual should desire, is what matters when it comes to how well-off the individual is. This understanding of individual utility enables KS to try to accommodate notions of Fairness, such as the notion that a situation that better approximates an ideal of distributive justice is a better situation.<sup>125</sup> Part II.B.2 discussed a different way that KS might try to accommodate distributive concerns—select the appropriate SWF. In this part, I focus on a type of preference that KS claim plays a significant role in their defense of welfare economics—a taste for Fairness.

KS assert that “[o]f particular relevance to [their] analysis is the possibility that individuals have tastes for legal rules that comport with some personally held notions of fairness.”<sup>126</sup> But why are tastes for Fairness particularly relevant to KS’s analysis? The main reason is that once we recognize that tastes for Fairness are only tastes, like a taste for chardonnay, we see that “satisfying the principle of fairness enhances the individual’s well-being, just as would satisfying his preference for wine.”<sup>127</sup> The idea is that welfarism, which focuses exclusively on individual well-being, can accommodate any Fairness notion as long as individuals prefer it.<sup>128</sup> The accommodation takes the form of caring about the utility of anyone who desires Fairness.<sup>129</sup> And, significantly, because welfarism makes room for Fairness in this way, intuitions that often motivate Fairness theorists become less attractive as reasons to eschew welfarism in favor of a Fairness view. If welfare economics can

---

individual’s preferences are not satisfied when the world is the way that she would like it to be but she does not believe this (and therefore is not particularly happy about it).

124. More precisely, what matters is what an individual actually *would* desire if she “he had all the relevant factual information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice.” John C. Harsanyi, *Morality and the Theory of Rational Behavior*, 44 SOC. RES. 623, 646 (1977). Such desires can be called an individual’s *true* desires, and according to welfare economics, satisfying a true desire always makes an individual better off, even if the content of the desire is objectionable (e.g., bigoted, sadistic). *Id.* I am inclined to disagree with the welfare-economic characterization of what makes an individual well-off, but I shall not challenge that characterization in this Article. I shall leave unchallenged the tendentious assumption that something is valuable precisely because it is desired (or would be under the circumstances that Harsanyi imagines), even independently of the reasons for which it is (would be) desired.

125. KAPLOW & SHAVELL, *FAIRNESS VERSUS WELFARE*, *supra* note 3, at 431–36.

126. *Id.* at 431.

127. *Id.* at 21.

128. *Id.*

129. *See id.* at 21–22.

consistently accommodate principles of Fairness, then what independent reason remains for being a Fairness theorist?

The representational capaciousness of desires plays a major role here:

[KS] adopt an expansive interpretation of the relevant preferences to be aggregated by the social welfare function. They pursue . . . a strategy of incorporation that reflects in the individual's preference ordering *every aspect of* a state of affairs that weighs in the individual's assessment of his actions. They assert that evaluation of legal decisions ought to depend only on this broad conception of preference . . . .<sup>130</sup>

Because welfare economics endorses a broad, preference-satisfaction interpretation of individual utility, welfare economists have available what might be thought of as a powerful "vacuum cleaner." Welfare economists can use this "representational" vacuum cleaner to "engulf" any possible feature of a situation within a welfare economics framework because such features might be represented by an individual's mental states, such as beliefs and desires. Welfare economics, in principle, cares about any notion insofar as the notion is an aspect of the content of someone's true tastes or preferences.<sup>131</sup> And crucially, "individuals [may] have a taste for a notion of fairness . . . . For example, an individual might derive pleasure from knowing . . . that legal rules reflect a favored conception of fairness."<sup>132</sup> In this way, welfare economics can try to accommodate Fairness features of situations indirectly, insofar as those features are reflected in the contents of individual preferences.<sup>133</sup> This incorporation strategy seems

130. Lewis A. Kornhauser, *Preference, Well-Being, and Morality in Social Decisions*, 32 J. LEGAL STUD. 303, 304 (2003) (emphasis added).

131. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 21.

132. *Id.* Along similar lines, "[j]ust as an individual might derive pleasure from art, nature, or fine wine, so might an individual feel better with the knowledge, for example, that vicious criminals receive their just deserts." Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3, at 333.

133. This way of trying to capture non-utility information within a framework that ultimately recognizes only utility information as relevant to a state's value is not original to KS. For example, J.S. Mill, who (roughly) equated utility with a subjectively experienced form of happiness, argued that a utilitarian could promote virtue for reasons other than any tendency for virtue, *as a means*, to bring about happiness:

The principle of utility does not mean that any given pleasure . . . [is] to be looked upon as means to a collective something termed happiness, and to be desired on that account. They are desired and desirable in and for themselves; besides being means,

to undermine any independent appeal a Fairness theory might have because KS can accommodate anything that the Fairness theorist might have to offer as long as enough people happen to prefer it.<sup>134</sup>

To evaluate KS's use of this incorporation strategy, it is important to recall precisely what KS claim Fairness *is*. KS offer the following explicit, stipulated definition:

By notions of fairness we include all principles—whether stated in terms of justice (such as corrective or retributive justice), rights (such as a right to a day in court), or cognate concepts (such as the sanctity of promises)—that may be employed to assess the desirability of policy and that have the following characteristic: *At least some weight is given to factors independent of individuals' well-being*. That is, we define notions of fairness to include all independent evaluative principles that are not purely welfarist.<sup>135</sup>

KS also remark that “most notions of fairness are nonconsequentialist” and that many Fairness-based recommendations concerning what society should do “do[] not depend exclusively (or at all) on an assessment of the consequences of doing these things.”<sup>136</sup> Thus, KS's definition of Fairness appears to include any notion that does *either or both* of the following: (a) when deciding what choice to take, place weight on something other than the consequences of the feasible options; (b) when evaluatively comparing two or more situations, place weight on features of those situations other than their utility information.

In the remainder of this part's discussion of tastes for Fairness, I shall focus on *only* those Fairness notions that place weight on something other than utility information when comparing how good one

---

they are a part of the end. Virtue, according to the utilitarian doctrine, is not naturally and originally part of the end, but it is capable of becoming so; and in those who live it disinterestedly it has become so, and is desired and cherished, not as a means to happiness, but as a part of their happiness.

J.S. MILL, UTILITARIANISM 36–37 (George Sher ed., Hackett Publ'g Co., 2d ed. 2001) (1861).

134. See KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 18–19.

135. Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3, at 333–34 (emphasis added). Along similar lines, KS state that they “employ the terminology of ‘notions of fairness’ to refer only to principles that accord weight to factors that are independent of individuals' well-being.” KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 44 (emphasis omitted).

136. Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3, at 334.

state is vis-à-vis another. In other words, for the remainder of this part, I take a notion of Fairness to be any notion that endorses (b). While this characterization includes only some of the views that KS would think fall under the heading of Fairness, KS's definition of Fairness commits them to agreeing that all views that endorse using Fairness SWFs to evaluate and compare situations are Fairness views.<sup>137</sup>

Having reemphasized what Fairness is, it is important next to turn to what KS think having a *taste for* Fairness is. As KS describe it:

if individuals in fact have tastes for notions of fairness—that is, *if they feel better off when laws that exist or events that they observe are in accord with what they consider to be fair*—then analysis under welfare economics will take such tastes into account when measuring individuals' well-being, just as it will take any other tastes into account.<sup>138</sup>

To take a specific example of such tastes for Fairness, Steven Shavell and A. Mitchell Polinsky have explained how we might account for desires that criminals get their just deserts within a welfare-economic analysis of optimal criminal sanctions.<sup>139</sup> They start with the standard welfare-economic deterrence model<sup>140</sup> and show how to incorporate a taste for “retributive” Fairness into it.<sup>141</sup> Of particular interest here is how they describe the taste that they attempt to accommodate:

To take beliefs concerning the fairness of sanctions into account, we assume that the utility of individuals depends on the sanctions that are imposed on those who violate laws—specifically, that individuals obtain . . . fairness-related utility or disutility from the imposition of sanctions on others. *Individuals might feel better off as a result of punishment of violators, perhaps due to retributive satisfaction, or individuals might feel worse off, because of empathy for violators.*<sup>142</sup>

---

137. See, e.g., KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24 n.15 & 38–39.

138. Kaplow & Shavell, *Fairness Versus Welfare*, *supra* note 3, at 1350 (emphasis added).

139. See A. Mitchell Polinsky & Steven Shavell, *The Fairness of Sanctions: Some Implications for Optimal Enforcement Policy*, 2 AM. L. & ECON. REV. 223 (2000).

140. See Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968), for a presentation of this basic model.

141. See Polinsky & Shavell, *supra* note 139, at 224.

142. See *id.* (emphasis omitted and added).

There are two points worth highlighting here. First, the foregoing characterizations of tastes, or preferences, for Fairness strongly suggest (consistent with this Article's working assumption about what satisfying a preference is) that an individual's preferences are satisfied only if that individual *believes* that the proposition (which is the content of her preference) has come true. There would be no plausible way to account for why an individual "*feels better*" or "*feels worse off*" if satisfying or dissatisfying her desires was independent of what she believed. Second, and significantly, it is not clear that the kinds of tastes described in the passages above are *for* Fairness if we understand Fairness to be defined the way KS carefully define that concept.

Given KS's technical definition of Fairness, to genuinely have a preference for Fairness it would seem that one must prefer that direct weight be placed on features of states other than their utility information when making comparative judgments about how good those states are (at least, when this is done by analysts engaging in critical-level thinking). Otherwise, it is not Fairness (in KS's sense) that is getting preferred. In other words, a taste for Fairness (in KS's sense) would seem to be a preference that a Fairness SWF be used to evaluate situations. If not, then whatever a taste for Fairness amounted to, it would not *be* a taste for the concept that KS carefully explicate. This poses a serious problem, however. By definition, a welfare-economic analyst could not accommodate such preferences if (1) accommodating such preferences amounted to satisfying them (for a preference cannot be "satisfied" without coming true) and (2) the analyst endorses (as KS do) a normative commitment regarding purely welfare-economic critical-level thinking by legal analysts, including herself. For to satisfy such tastes, the analyst *must* use a Fairness SWF to evaluate situations.

There are at least two ways KS might try to avoid this problem. The first would be to acknowledge that when they discuss tastes for Fairness, KS do not literally mean a taste for Fairness in the sense that they technically define. Consider one description of what such a taste (say, one of Linda's) is *for*—a taste for legal doctrines that accord with what Linda considers to be "fair." What Linda has in mind by fair and the basis on which she would determine whether a given law was fair are unclear here. For example, KS acknowledge that what Linda thinks of as fair might have nothing to do with Fairness as KS technically define it but instead turn on what Linda might think best for society:

Another difficulty in ascertaining individuals' tastes is that their opinions concerning which legal rules and procedures are fair may reflect their views about which schemes are best for

society rather than a direct taste, say, for a particular legal doctrine. For example, an individual who thinks that current medical malpractice liability is unfairly onerous may not feel worse off directly as a result of the rule's existence but may merely believe that the current regime is unwise because it raises medical costs.<sup>143</sup>

The “tastes for fairness” that welfare economics can accommodate generally appear to be tastes whose satisfaction makes an individual feel good in some way. An example would be a taste that criminals be severely punished, a taste whose satisfaction makes the individual feel, so to speak, retributively gratified. Retributive *reasons*, as opposed to vengeance-related “moral” emotions, do not seem central to the sorts of tastes for fairness that welfare economics can actually accommodate.

Having considered the first possible way out of the difficulty—conceding that the sorts of tastes for fairness that welfare economists can endorse satisfying are not really tastes for Fairness in KS's official, technical sense—I now turn to a second. Perhaps a welfare economist could insist that he can *accommodate* a taste for the version of the notion of Fairness that KS carefully define. The welfare economist might concede that it is impossible for him consistently to endorse actually *satisfying* any taste for Fairness. But then he could claim that welfare economics can *accommodate* tastes for Fairness without satisfying them. What the welfare economist can endorse is satisfying various “moral-emotionally driven” tastes that *correspond* in the right way to genuine notions of Fairness. For *if* these gratified moral emotions correspond in the “right way” to genuine tastes for Fairness, *then* by satisfying the former, the latter can be *accommodated*.

Although, to my knowledge, KS never distinguish “satisfying” a preference and “accommodating” it, they do contemplate the idea that social norms, whose adoption might somehow reflect moral emotions, *correspond to* notions of Fairness:

---

143. Kaplow & Shavell, *Fairness Versus Welfare*, *supra* note 3, at 1351–52 (emphasis added). In passing, I should remark that KS's mention of a “direct taste” for a legal doctrine here is tantalizing because it suggests that *if* Linda had a direct taste for a particular legal doctrine instead of forming a taste for it because she thought satisfying it would be best for society, *then* the taste would be a taste for Fairness of the sort KS have in mind. But this route seems closed to KS because, as discussed above, KS are explicit that the *reason why* Linda happens to have whatever tastes she has is unimportant. It would therefore be of no particular importance that Linda has a *direct* taste for Fairness that might be satisfied. Also, it is not clear exactly what KS mean by direct here, unless they mean something like “preferred for no independent reason.”

We submit that there is often a correspondence—indeed, sometimes an identity—between notions of fairness that are used as independent principles for the evaluation of legal rules and various social norms that guide ordinary individuals in their everyday lives. Moreover, we suggest that this relationship between notions of fairness and social norms helps to reconcile the attraction that notions of fairness possess with our argument that such notions should not be given independent weight in the assessment of legal rules.<sup>144</sup>

Developing this line, KS could argue that the moral-emotionally driven tastes that welfare economists can endorse satisfying are tastes that various social norms in some sense be respected.<sup>145</sup> Then, KS could try to argue that by satisfying such tastes, welfare economics could *accommodate* tastes for the notions of Fairness (now in the technical sense) to which those social norms correspond. But to do this KS would need to abandon the idea that there can ever be an *identity* between such social norms and the corresponding notions of Fairness. For if there were such an identity, then satisfying a taste for respecting the social norm would *just be* satisfying (not merely accommodating) a taste for Fairness. For the second way out of the dilemma to work, KS would need to explain how to accommodate genuine tastes for Fairness without satisfying them, for it is contradictory for KS to endorse *satisfying* such tastes (assuming that to “satisfy” a taste one must at least make it come true).

Regardless of which of the two paths KS take, by the time they have responded to the challenge, it is unclear why welfare economics’ ability to accommodate tastes for Fairness would be especially significant. KS describe satisfying tastes for Fairness as a “particular source of well-being that has special relevance”<sup>146</sup> to their project. To reiterate one possible way this source of well-being *could be* of special relevance: *If* welfare economics can accommodate tastes for Fairness, *then* Fairness theorists are left without independent reasons to resist welfare economics. Welfare economists could just turn on their representational vacuum cleaner and engulf any putative Fairness-based reason to resist

---

144. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 62–63.

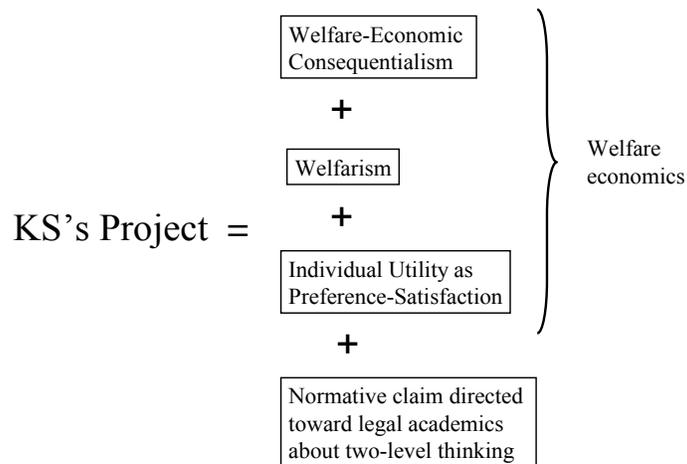
145. The tough part would be for KS to explain in what sense such norms can be “respected” without placing any direct weight on those norms as *reasons* when assessing policy options. For if direct weight is placed on those norms, then welfare economics is violated.

146. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 21.

welfare economics into welfare economics itself. But if what welfare economics can endorse satisfying are not really tastes for Fairness (as KS define Fairness), then there is no reason to think that what gets engulfed *are* the Fairness-based reasons to resist welfare economics. And if welfare economics cannot endorse satisfying preferences for Fairness but can only endorse “accommodating” them in some weaker sense, then there is no reason to think that the Fairness-based reasons to resist welfare economics actually get *engulfed*—those reasons are still “out there,” untouched by the representational vacuum cleaner.

#### 4. Recapitulating KS’s Project

In light of the foregoing, the main substantive elements of KS’s project can be diagrammatically summarized:



And corresponding to the four main elements of KS’s project are four propositions, which KS must support for their project to succeed:

- (1) Welfare-economic consequentialism is true.
- (2) Welfarism is true.
- (3) The “preference-satisfaction” account of individual well-being is true.
- (4) Legal and policy analysts should engage exclusively in welfare-economic critical-level thinking when generating and justifying normative (i.e., prescriptive or proscriptive) legal and policy scholarship.

## III. KAPLOW AND SHAVELL'S PARETO ARGUMENT

## A. Preliminaries—Pareto Efficiency and the Pareto Principle

To support (1)–(4) KS offer many significant arguments, observations, definitions, explications, and illustrations.<sup>147</sup> One of these is KS's Pareto argument, which attracted considerable scholarly attention. KS's Pareto argument is deductively valid.<sup>148</sup> Yet in Part IV, I shall raise a criticism: Even if deductively *sound*,<sup>149</sup> KS's demonstration provides no independent reason to believe *any one* of (1)–(4). As I shall explain, KS never thought that their Pareto argument provided independent reasons to believe, let alone demonstrated, *every one* of (1)–(4). But it would be problematic if KS's Pareto argument provided no independent (non-circular) reason whatsoever to accept *any* part of their project. I am not suggesting that this alone would show that welfare economics is false, or even that KS's normative thesis about critical-level thinking is false. However, KS think of their Pareto argument as very important, and they seem to place at least some reliance on it to further their project. And if this Article's analysis is correct, then KS's Pareto demonstration should not even be considered an argument in their project's favor.

Consider a principle: *If* an argument provides absolutely no independent reason to accept any part of a project, *then* that argument cannot properly be considered an argument in favor of that project. My criticism boils down to applying that principle to the relationship between KS's Pareto demonstration and their project. In Part IV, I shall explain why KS's Pareto demonstration (a) provides no independent reason to endorse any part of welfare economics, and (b) provides no independent reason to think that legal academics should engage purely in welfare-economic critical-level thinking when generating, promulgating, and justifying normative advice about how to address important social choices.

KS's Pareto demonstration draws upon a well-known notion of economic efficiency—Pareto efficiency: “[I]f we can find a way to make some people better off without hurting anybody else, then we have an

---

147. See generally *supra* note 3.

148. A deductive argument is valid when its conclusion follows of logical necessity from its premises—that is, when it is logically impossible that its premises be true and its conclusion false.

149. A deductive argument is sound when (i) it is valid, and (ii) all of its premises are true.

allocation that is Pareto inefficient. If no such Pareto improvement can be found, the allocation is Pareto efficient.”<sup>150</sup> Notice up front that Pareto efficiency depends on feasibility. Our ability or inability to find a way to make some better off and none worse off is a function in part of what we feasibly can do when we are trying to figure out whether a situation is Pareto efficient. Whether a situation is Pareto efficient is therefore in part a function of highly contingent factors such as available technology, resources, etc. For example, an allocation of utility in a population that may have rendered a situation Pareto efficient in 1850 may well not do so in 2100, just in virtue of changes in technological know-how and access to resources between 1850 and 2100.

Pareto efficiency suggests what may seem to be a very plausible principle regarding the comparative values of two states: If at least one person’s utility is higher in state A than in state B and if no one’s utility is lower in state A than in state B, then A is better than B. The version of this principle operative in KS’s Pareto argument may seem even less controversial:

(P) If everyone’s utility is higher by some amount (not necessarily the same per person) in state A than in state B, then A is better than B.<sup>151</sup>

(P) is commonly referred to as the weak Pareto principle.<sup>152</sup> Notice that, unlike the notion of efficiency related to it, the weak Pareto principle has nothing (*per se*) to do with feasibility. Indeed, (P) says nothing about whether the states being compared are (or even could

150. HAL R. VARIAN, *INTERMEDIATE MICROECONOMICS* 15 (4th ed. 1996). This definition refers to “allocations” as the sorts of things that can be Pareto efficient. But note that an allocation of goods can be thought of as a property of a situation. So to enhance generality, I would (consistently with KS) use the term “Pareto efficient” to describe situations, states of affairs, or outcomes.

151. Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283.

152. This terminology has the potential to confuse. A different maxim—of importance in business management contexts and known as the “80/20 Rule,” among other monikers—also goes by the name “Pareto Principle” or “Pareto’s Principle”: “[A] small number of causes is responsible for a large percentage of the effect, in a ratio of about 20:80.” *Pareto Principal*, WIKIPEDIA (March 3, 2016), [http://en.wikipedia.org/wiki/Pareto\\_principle-cite\\_note-aka-0](http://en.wikipedia.org/wiki/Pareto_principle-cite_note-aka-0) [https://perma.cc/PCE3-5Q7A]; Arthur W. Hafner, *Pareto’s Principle: The 80-20 Rule*, BALL ST. U. (Mar. 31, 2001), <http://www.bsu.edu/libraries/ahafner/awh-th-math-pareto.html> [https://perma.cc/QKY8-V46P]. Note also that one could think of the weak Pareto principle as “more directly” suggested by a corresponding version of Pareto efficiency, according to which (i) if there is a way to make everyone better off, then our current state is Pareto inefficient, and (ii) if there is no such way, then our current state is Pareto efficient.

ever realistically be) the outcomes of legal or policy choices that are feasible for an agent (whether that agent is an individual or a group). Of course, welfare economists care a lot about what actually is or is not feasible, so the implications of the weak Pareto principle might become especially important when the states being compared could actually be brought about by implementing a legal change or a new policy. But it is significant that the weak Pareto principle is not about what actually can or cannot be brought about. The weak Pareto principle is not, strictly speaking, about the assessment of policies, but is instead about the assessment of states or situations. And the states being comparatively assessed could be completely fictitious.<sup>153</sup>

KS sometimes discuss their Pareto argument in connection with two types of settings, or contexts—a symmetric setting and an asymmetric setting.<sup>154</sup> A symmetric setting is an assumed “starting position” in which “everyone has the same income and preferences, is exposed to the same risks, and has the same opportunities.”<sup>155</sup> As an example of a symmetric setting that is useful when considering how to deal with automobile accidents, one could assume that *ex ante* everyone is as likely to be an injurer as a victim; in a contracts scenario, one could assume that *ex ante* everyone is as likely to breach a contract as to “suffer” from a breach; and so forth. Because, for KS, an individual’s utility depends only on satisfying that individual’s preferences, any social choice made in a symmetric situation affects the utility of every individual identically, at least on average.<sup>156</sup> In symmetric contexts, every individual has the same preferences concerning whatever issue the analyst is considering, and the probability that those preferences become satisfied is also the same because the income, opportunities, and risks of all individuals are also identical.<sup>157</sup> There is a temporal sequence: (1) symmetric starting position; then (2) policy decision (rule

---

153. Even if we were comparing situations that were not just fictional but fantastical, the weak Pareto principle would constrain comparative evaluations. To make this graphic, imagine that the only bearers of utility in state A are twenty-five purple polka-dotted, fifty-foot-long, sentient, desire-having alien slugs from “Planet X.” And imagine that exactly those same utility bearers are in state B. Assume that all of them are worse off in state B than in state A. (P) would *require* that state A is more valuable (“better than,” “manifests more ethical goodness than”) state B.

154. See, e.g., Kaplow & Shavell, *The Conflict Between Notions of Fairness*, *supra* note 3, at 68.

155. *Id.* at 68 n.11.

156. See *id.* at 68.

157. *Id.* at 68 & n.11.

choice); and then (3) resulting state.<sup>158</sup> To reemphasize, the expected change in utility between the starting position and the resulting state is the same for each individual. In this way, symmetric settings assure (on average) identical effects of policy choices upon individuals. “If everyone in society is [on average] identically affected by the rule, then either the rule will [on average] make everyone better off or it will [on average] make everyone worse off.”<sup>159</sup>

In contrast, in an asymmetric setting not everyone is similarly situated. If one is considering automobile accidents, for example, one could assume that some people drive and some do not. So a social choice in an asymmetric context may (and most likely will) affect the utilities of individuals differently. To reiterate one important point: Any context or setting (symmetric or asymmetric) is the first element of a temporally-extended structure: (1) a starting situation; then (2) a choice; and then (3) an ending situation (which the choice brings about). This structure is presupposed by the point of distinguishing symmetric from asymmetric contexts to begin with.

I shall focus on KS’s Pareto argument as applied generally—that is, without assuming a symmetric context. If KS’s Pareto argument works in an asymmetric setting, then it will work in a symmetric setting, too, as will become apparent when I turn to KS’s demonstration (as I will very shortly).<sup>160</sup> In fact, the temporal sequence referred to above (starting position; rule choice; resulting state), which is required to make sense of drawing the distinction between symmetric and asymmetric settings to begin with, does not play *any* role in KS’s Pareto argument.<sup>161</sup> That said, however, when welfarism is combined with consequentialism (or, for that matter, any plausible non-consequentialist view that takes the

158. *Id.* at 68.

159. Craswell, *supra* note 7, at 246. And KS point out that when one deviates from welfare economics, everyone actually becomes worse off in such situations:

[I]n symmetric contexts . . . it is always the case that everyone will be worse off when a notion of fairness leads to the choice of a different legal rule from that chosen under welfare economics. . . . Because everyone is identically situated, whenever welfare economics leads to the choice of one rule over another, it must be that everyone is better off under the preferred rule. Hence, whenever a notion of fairness leads one to choose a different rule from that favored under welfare economics, everyone is worse off as a result.

KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 52 (emphasis omitted).

160. *See infra* Part III.B.

161. *See generally* Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283–85.

consequences of feasible options seriously), the symmetric/asymmetric distinction between contexts becomes potentially important. And, as KS point out, symmetric contexts can provide particularly vivid examples of the potential superiority of welfare economics to other normative systems.<sup>162</sup>

### B. *The Pareto Argument*

In reconstructing KS's Pareto argument, I shall very closely draw upon on the version in a brief 2001 article by KS published in *The Journal of Political Economy*.<sup>163</sup> I take that version to be KS's "official" version, a version that makes clear exactly what KS's premises are, what their inferences are, and what their conclusion is. In other of their writings, KS offer heuristic, "looser" versions of the Pareto argument, but to assure addressing their actual argument, technical assumptions and all, I shall focus on *The Journal of Political Economy* version. As mentioned previously, the Pareto argument is a deductive demonstration. This demonstration shows that any Fairness SWF (that satisfies certain assumptions) violates a weak version of the Pareto principle:

(P) If everyone's utility is higher by some amount (not necessarily the same per person) in state A than in state B, then A is better than B.<sup>164</sup>

As previously explained, KS provide a formal specification of the type of SWF consistent with welfarism:  $W(x) = F(U_1(x))$ ,

---

162.

[O]ur demonstration that all notions of fairness always make everyone worse off in the symmetric case is especially significant under many broadly endorsed normative frameworks. We show that the Golden Rule, Kant's categorical imperative, and the veil-of-ignorance construct each require that normative principles be tested as if one is in a symmetric setting. . . . Accordingly, if one adheres to any or all of these normative frameworks—as most fairness proponents in fact do—one is forced either to reject all . . . principles [that deviate from welfare economics] or to endorse the view that making everyone worse off should be the core feature of any sound normative principle.

Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3, at 337 (citation omitted).

163. Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3.

164. Note that (P) is supposed to be *necessarily* true. In other words, according to (P), it is *logically impossible* for a situation in which everyone is worse off to be better than one in which everyone is better off.

$U_2(x), \dots, U_n(x)$ ).<sup>165</sup> (Here,  $U_i(x)$  is the utility function for individual  $i$  from a society comprising  $n$  individuals and  $x$  ranges over states of affairs.)<sup>166</sup> In contrast, a Fairness SWF places direct weight on a feature of a situation other than that situation's utility information.<sup>167</sup> Accordingly, a Fairness SWF can be represented as:  $Z(x) = F(U_1(x), U_2(x), \dots, U_n(x), f(x))$ ,<sup>168</sup> where  $f(x)$  is a function that assigns to any state  $x$  a number representing the amount of Fairness in  $x$ . Here,  $x$ 's Fairness consists of all of  $x$ 's features (*other than*  $x$ 's utility information) that  $Z$  considers relevant to  $x$ 's value. Thus,  $f(x)$  is a function that assigns to any state  $x$  a real number representing the amount of Fairness in  $x$  based directly on whether  $x$  has those features, not on how those features get "filtered through" anyone's utility function. (A terminological point: The domain of any individual utility function or social welfare function can be represented as  $X$ , which stands for all possible situations. Thus,  $x \in X$  simply means that  $x$  is a situation, or state of affairs, which may or may not be actual.)

KS correctly assert, in what they term an "OBSERVATION," that for any  $Z(x)$ , there must exist  $x, y \in X$  such that  $U_n(x) = U_n(y)$  for all  $n$  and  $Z(x) \neq Z(y)$ .<sup>169</sup> That is, *any* Fairness SWF must evaluate at least two logically possible states (containing the same individuals with the same utilities) differently. The reason is that if there were no such states—if it

165. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24 n.15.

166. *See id.*

167. *See id.*

168. KS give Fairness SWFs the following form:  $Z(x) = F(U_1(x), U_2(x), \dots, U_n(x), x)$ . *See id.* at 39–40 n.52. I modify the last component of the function's input to  $f(x)$ . KS want to emphasize how features of the states being evaluated are directly factored into the Fairness SWF. Hence, they leave  $x$ , which ranges over propositions or states, naked in the last component of the function's input—for a non-welfarist,  $x$  need not affect a SWF's value as an argument of an individual utility function. I think that it is also helpful to emphasize that a state's features "directly" influence  $Z$ 's value "via" a real number (or perhaps an ordered n-tuple of real numbers), representing the amount(s) of (irreducibly different types of) Fairness (non-utility information) regarding that state. In KS's official version of the Pareto argument, Fairness comes in amounts, just as utility does. Strictly speaking, the "naked" variable  $x$  does not take numbers as possible values because  $x$  ranges over situations, not numbers. The  $U_n(x)$  convert states into numbers representing individual utilities in those states. I introduce  $f(x)$  to do something analogous vis-à-vis Fairness:  $f(x)$  converts a situation into a real number (or ordered n-tuple of real numbers) representing how much (of each type of) Fairness that situation has. I have no substantive disagreement with KS here—my variation of their form simply modifies what gets emphasized.

169. *See* Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283. Following is the "OBSERVATION," as KS express it: "A social welfare function  $F$  is not individualistic if and only if there exist  $x, x' \in X$  such that  $U_i(x) = U_i(x')$  for all  $i$  and  $F(x) \neq F(x')$ ." *Id.* at 283.

were *impossible* for anything but utility information *ever* to make a difference to how a SWF evaluated a situation—then that SWF would be of the welfarist form  $W(x)$ , devoid of even an iota of Fairness, as KS define that concept. Call this observation (O). With (O) in hand, KS proceed as follows. They make two further assumptions:

(A1) It is possible that each person have more of some good, such that if each person did have more (by even a tiny amount), then each person's utility would increase by some finite (perhaps tiny) amount.<sup>170</sup> In other words, for any possible state  $x$  and for any specified amount of the good, there is another possible state  $x'$  such that:  $x'$  is just like  $x$  except that everyone in  $x'$  has that amount more of the good than in  $x$ ; and everyone's utility is thereby higher in  $x'$  than in  $x$ .<sup>171</sup>

(A2) Any Fairness SWF,  $Z$ , is continuous on the extra amount of the good that each individual has. In other words, as one continuously increases the amount of the good that each person has, the state improves (under  $Z$ ) incrementally, without any jarring, dramatic increases.<sup>172</sup>

Now take *any* Fairness SWF,  $Z(x)$ , that satisfies KS's assumptions. By (O), there must exist two distinct possible states (call them  $W$  and  $V$ ) such that (1) both  $W$  and  $V$  have the same individuals with the same utility function values, and (2)  $Z(W) \neq Z(V)$ . (The difference between  $Z(W)$  and  $Z(V)$  would be attributable to differing amounts of Fairness

170. To clarify, the small additional amount of the good that each person (say, Bill and Jill) gets is the same, but the amount by which Bill's utility increases might not be the same as the amount by which Jill's does. Also, we are not assuming that Bill and Jill have any of the good before the increase, though they might "start off with" some before they get their additional amount. It is important, however, that the utility of both of them increases as they get more and more of the good.

171. Following is (A1), as KS express it:

There exists a good such that, if each person has  $\delta$  more of it, then each person is better off. Specifically, let  $m_i$  be individual  $i$ 's amount of the good. Then if two states,  $x$  and  $x'$ , are identical except that, for all  $i$ ,  $m_i$  in  $x$  is higher by  $\delta > 0$  than  $m_i$  in  $x'$ , then  $U_i(x) > U_i(x')$  for all  $i$ .

Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 283–84.

172. Following is (A2), as KS express it (here, ' $F$ ' stands for a social welfare function): "The function  $F$  is continuous in the  $m_i$ ." *Id.* at 284. Note that some have pointed out that KS's need to assume the continuity of  $Z$  severely restricts the scope of KS's demonstration. Such critics have argued that plausible Fairness SWFs that are discontinuous at points where the Pareto principle might otherwise be violated escape KS's demonstration. *See, e.g.*, Chang, *supra* note 7; Craswell, *supra* note 7. KS are unmoved by this type of argument. *See* Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3.

in each state (i.e.,  $f(W) \neq f(V)$ .) Also, assume that  $Z$  assigns the larger real number to  $V$ , so that  $Z(V) > Z(W)$ .<sup>173</sup> Thus,  $Z(V) - Z(W) = r$ , some positive real number. Now consider a possible state that is just like  $W$ , except that everyone has a little bit more of the good discussed earlier. By (A1), we know that there is such a logically possible state, which we can call  $W'$ . Also by (A1), everyone's utility is higher in  $W'$  than in  $W$ . By (A2), as we incrementally and continuously increase the amount of the good that everyone in  $W$  gets,  $Z(W')$  incrementally and continuously gets higher, too, without any "gaps" or "jumps." Metaphorically, as we gradually "turn up the dial" on how much more of the good each person gets in  $W$ , we gradually turn up the dial on how highly  $Z$  values the resulting state  $W'$ .

When "mathematically" constructing  $W'$ , imagine that we increase the extra amount of the good everyone gets in  $W$  just enough so that, even though  $Z(W') > Z(W)$ ,  $Z(W')$  is not too much larger. Specifically, we make sure that  $0 < (Z(W') - Z(W)) < r$ . Since  $Z(V) - Z(W) = r$ , substituting for  $r$  yields  $Z(W') - Z(W) < Z(V) - Z(W)$ . Thus,  $Z(V) > Z(W') > Z(W)$ . So the Fairness SWF evaluates  $V$  as better than  $W'$ . Recall that everyone's utility in  $V$  is the same as it is in  $W$ . Because everyone's utility is higher in  $W'$  than in  $W$  (because of the increase in the good), it follows that everyone's utility is also higher in  $W'$  than in  $V$ . Because  $Z$  evaluates  $V$  as better than  $W'$ ,  $Z$  violates (P), the weak Pareto principle. But  $Z$  was *any* Fairness SWF for which (O), (A1), and (A2) are true. Therefore, *any* such Fairness SWF violates the weak Pareto principle, and the proof is complete. In a nutshell, as long as one (1) assumes the logical possibility of<sup>174</sup> a good satisfying (A1) and (A2), and (2) limits one's attention only to Fairness SWFs satisfying (O), (A1), and (A2), it follows that no such Fairness SWF could always satisfy (P).

#### IV. KS'S PARETO ARGUMENT DOES NOT FURTHER THEIR PROJECT

KS's demonstration is deductively valid. If one grants (O), (A1), and (A2), then the conclusion necessarily follows. On those three assumptions, any Fairness SWF that is continuous<sup>175</sup> must violate (P) for

---

173. If we assume instead that  $Z(W) > Z(V)$ , the proof goes through just the same, *mutatis mutandis*.

174. To emphasize, the good need not be realistic; it just needs to be coherently conceivable.

175. More specifically, the SWF must be continuous on the amount of the good that each individual has when we construct  $W'$ . (A2) requires that the SWF be continuous in this

at least one pair-wise evaluation of two logically possible states. (We called those states V and W'.) Now a crucial question arises: *What are the implications of this result for KS's project?* KS claim that their Pareto argument bolsters their project.<sup>176</sup> For example, early in their book *Fairness Versus Welfare*, in which they carry out the lion's share of their project, KS include a section in which they offer what they call an overview of their main argument in the book.<sup>177</sup> In that section, KS state:

Our argument for basing the evaluation of legal rules entirely on welfare economics, giving no weight to notions of fairness, derives from the fundamental characteristic of fairness-based assessment: Such assessment does not depend exclusively on the effects of legal rules on individuals' well-being. As a consequence, satisfying notions of fairness can make individuals worse off, that is, reduce social welfare. Furthermore, individuals will be made worse off overall whenever consideration of fairness leads to the choice of a regime different from that which would be adopted under welfare economics because, by definition, the two approaches conflict when a regime with greater overall well-being is rejected on grounds of fairness.

This point takes on special force when, as we show in important situations [symmetric contexts], fairness-based analysis leads to the choice of legal rules that reduce the well-being of *every* individual. . . .

Indeed, the possibility that pursuing a notion of fairness may make everyone worse off is always present. . . . It can be demonstrated that consistently adhering to any notion of fairness will sometimes entail favoring regimes under which every person is made worse off.<sup>178</sup>

Here, KS refer to their Pareto result as if it were a central aspect of their main argument that legal analysis should be entirely welfare-economic.

As mentioned earlier, KS's project can be broken into four elements, to each of which corresponds a claim that KS must convincingly defend if their project is to succeed:

---

respect.

176. KAPLOW & SHAVELL, *FAIRNESS VERSUS WELFARE*, *supra* note 3, at 56.

177. *Id.* at 52–62.

178. *Id.* at 52–53 (footnotes omitted).

- (1) Welfare-economic consequentialism is true.
- (2) Welfarism is true.
- (3) The “preference-satisfaction” account of individual well-being is true.
- (4) Legal and policy analysts should engage exclusively in welfare-economic critical-level thinking when generating and justifying normative (i.e., prescriptive or proscriptive) legal and policy scholarship.

It bears reemphasis that defending (1), (2), and (3) is crucial to defending (4) because (4) is a strong normative claim about legal academics that makes reference, not just to any sort of critical-level thinking, but to *welfare-economic* critical-level thinking. Welfare economics are constituted by (1)–(3), so an inability to defend (1)–(3) amounts to an inability to defend welfare economics. And if one cannot defend welfare economics, then it is hard to see how one could hope to defend the very strong claim that if a legal or policy analyst ever deviates at all from welfare economics in addressing a normative issue, then the analyst has not done her job the way that she *should* have done it. If even one of the moral criteria of welfare economics is incorrect, then what convincing reason could there be always to think in terms of welfare economics to the exclusion of any other possible moral theory? In the remainder of this part, I shall explain why KS’s Pareto argument provides absolutely no non-circular (i.e., non-question-begging) reason to believe (1), (2), (3), or (4).

A. *KS’s Pareto Argument Provides No Reason to Believe (1)*

What about (1), welfare-economic consequentialism? What reason, if any, does KS’s Pareto argument give us to embrace consequentialism? Just before executing their formal derivation, KS state:

We show that, for any non-welfarist method of *policy assessment* (i.e., for any nonindividualistic social welfare function), there always exist circumstances in which the Pareto principle is violated. In other words, any conceivable notion of social welfare that does not depend solely on individuals’ utilities will sometimes *require adoption of a policy* that makes every person worse off.<sup>179</sup>

---

179. Kaplow & Shavell, *Any Non-Welfarist Method*, *supra* note 3, at 282 (emphasis added).

Right after executing their formal derivation KS declare:

We have demonstrated that any method of *policy assessment* that is not purely welfarist violates the Pareto principle. That is, policy evaluation that gives any weight to principles independently of their *effect* on individuals' utilities will sometimes lead to choices under which everyone is worse off.<sup>180</sup>

These claims about what KS's Pareto argument demonstrates are false. It is difficult to see why without sharply distinguishing welfarism from consequentialism. As previously explained, welfarism is not a species of consequentialism.<sup>181</sup> Welfarism *could be* combined with consequentialism, as it is by welfare economics.<sup>182</sup> But it need not be. Welfarism is a view about which features of a situation can affect that situation's value.<sup>183</sup> Specifically, it is the view that the only features of a state that determine the state's value are, collectively, the state's utility information.<sup>184</sup> Welfarism itself, in other words, constrains comparative assessments of *states* or *situations*—not assessments of *policies*. Maybe the states being compared would be the outcomes of different feasible policy choices that would affect the well-being of individuals, and maybe they would not. The states may be within our reach to actualize, or they may be entirely fictitious (perhaps even fantastical) and unattainable. Yet KS claim to *demonstrate* that the weak Pareto principle (P) is violated for any non-welfarist method of *policy* assessment, where the policies being assessed would be affecting the well-being of individuals. The weak Pareto principle does not have anything to do with policies or the effects of policies. It is therefore impossible for a method of *policy* assessment to violate (P), and thus impossible to demonstrate that (P) can be so violated. Recall (P):

(P) If everyone's utility is higher by some amount (not necessarily the same per person) in state A than in state B, then A is better than B.<sup>185</sup>

---

180. *Id.* at 284 (emphasis added).

181. *See supra* note 80 and accompanying text.

182. *See supra* Part II.B.2.

183. *See, e.g.*, KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 24 n.15.

184. *See supra* Part II.B.2.

185. By way of reminder, (P) is supposed to be *necessarily* true. In other words, according to (P), it is *logically impossible* for a situation in which everyone is worse off to be better than one in which everyone is better off.

A SWF would violate (P) if it ranked B as better than A. But close inspection of (P) reveals that (P) in itself says nothing about how *policies* are assessed. The only way that (P) could imply anything about how policies are to be assessed is if it is also *assumed* that assessing states is equivalent to assessing policies. As explained in Part II.B.1, these two types of assessments are far from equivalent,<sup>186</sup> so further (very big) implicit assumptions would be required to, as it were, “force” them into equivalence. One assumption that would help do this is the assumption that consequentialism is true. *If* consequentialism were true, then, arguably, assessing states would yield an assessment of policies because on that assumption what makes implementing a policy permissible, impermissible, and so forth is the value of that policy’s outcome (a resulting state), as compared with the values of the outcomes of competing feasible policy choices. But in this context, that is an *enormous* “if.” It is enormous because what is currently at issue is whether the Pareto demonstration provides even the slightest reason to think that consequentialism is true. If it is assumed that consequentialism is true, then anything demonstrates that consequentialism is true.<sup>187</sup>

Further, KS’s Pareto demonstration has only three premises: (O), (A1), and (A2). Re-inspecting (O), (A1), and (A2) verifies that the key concepts of consequentialism (including feasibility, permissibility, and maximization)<sup>188</sup> are entirely absent:

(O) For any  $Z(x)$ , there must exist  $x, y \in X$  such that  $U_n(x) = U_n(y)$  for all  $n$  and  $Z(x) \neq Z(y)$ .

(A1) It is possible that each person have more of some good, such that if each person did have more (by even a tiny amount), then each person’s utility would increase by some finite (perhaps tiny) amount. In other words, for any possible state  $x$  and for any specified amount of the good, there is another possible state

186. Recall that consequentialism cannot, for example, handle positional reasons that might bear on whether a particular policy choice is permissible.

187. To bring this out vividly, consequentialism follows deductively from the following two assumptions: (1) consequentialism is true, and (2) Jackie Gleason was a prime number. Whatever one plugs in to (2), *anything* can be added as a part of a “demonstration” that consequentialism is true if one *assumes* that consequentialism is true.

188. As explained, according to consequentialism only the consequences of implementing feasible options are relevant to what choice morally ought to be made, and for a feasible action to be morally permissible, the consequence that the action brings about must be no worse than the outcome of any other feasible action.

$x'$  such that  $x'$  is just like  $x$  except that everyone in  $x'$  has that amount more of the good than in  $x$  and everyone's utility is thereby higher in  $x'$  than in  $x$ .

(A2) Any Fairness SWF,  $Z$ , is continuous on the extra amount of the good each that individual has. In other words, as one continuously increases the amount of the good that each person has, the state improves (under  $Z$ ) incrementally, without any jarring, dramatic increases.

In short, KS's demonstration provides absolutely no reason to think that consequentialism is true. Pointing this out does not, of course, amount to a reason to think that consequentialism is false. To be clear, I am not arguing at this point that welfare-economic consequentialism is false. I am just arguing that KS's Pareto demonstration provides absolutely no reason to think that consequentialism is true.

Related to the disconnect between KS's Pareto argument and consequentialism is KS's claim to have demonstrated that any Fairness SWF sometimes *requires the adoption of a policy* that will make everyone worse off. In what sense, exactly, would a Fairness agent (i.e., an agent who denies welfarism and evaluates states with a Fairness SWF) be required by her own view (even sometimes) to adopt a policy that will make everyone worse off? The Pareto argument does not demonstrate this. Remember that in KS's demonstration everyone is worse off in  $V$  than in  $W'$  and the Fairness SWF indicates that  $V$  is better than  $W'$ . So, to be sure, the Fairness agent, logically, must think that  $V$  (where everyone is worse off) would be a better way for the world to be than  $W'$  (where everyone is better off). But this demonstrates nothing about what the Fairness agent is logically committed to thinking as to what policy she is morally required to implement when facing any social choice. For KS's demonstration to show that, it would need to be significantly altered. For example, KS could add the following additional (very substantive) assumptions: (a) the Fairness agent happens to be a consequentialist, and (b) the Fairness agent could be faced with a forced social choice between exactly two feasible policies, one that would bring about  $V$  and one that would bring about  $W'$ . *If* KS add these assumptions, *then* the new version of the demonstration would show that the Fairness theorist is logically committed by her own view to the conclusion that she sometimes is required to choose to implement a policy that makes everyone worse off. But what price would KS need to pay to achieve this result in this way?

KS must add assumption (a) to get this result because, if the Fairness agent is not a consequentialist, then the agent may well not be logically committed by her own view to thinking that she is required to implement the policy that makes everyone worse off, even if she were faced with the choice of bringing about V or bringing about W'. If a non-consequentialist agent were faced with *that* choice, her theory would *require* her to consider at least some reason other than the comparative value of V versus W' to be relevant to what she would be permitted to do. That reason may, or may not, (depending on the details of the Fairness agent's non-consequentialist theory) outweigh or trump V's being better than W'. The non-consequentialist Fairness agent *might* be logically required by her theory to think, therefore, that the only morally permissible policy choice is the one that brings about W', thereby having the "side-effect" of making everyone better off. Perhaps ironically, the non-consequentialist Fairness agent might in some sense regret that she was required by her theory to do what made everyone better off at the cost of making the world a worse place than it would have been if she had been permitted to bring about the better state, V (in which everyone is worse off).

That KS must add assumption (a) would be problematic for them. KS are well aware that there are two different ways of being a Fairness theorist: (1) deny consequentialism, or (2) embrace consequentialism while denying welfarism, claiming that some feature of a state other than the state's utility information contributes to that state's value.

We will not usually distinguish between notions of fairness that do not depend on any consequences and those that do depend on consequences but not on individuals' well-being. (It is, of course, logically possible to have consequentialist principles that are independent of welfare . . . .) Because our claim is that consequences for individuals' well-being are what should count in evaluation, not consequences unrelated to their well-being, we do not see a consequentialist fairness principle that is unrelated to well-being as better than a nonconsequentialist principle.<sup>189</sup>

KS also aver (I think correctly) that most Fairness theories are non-consequentialist:

[M]ost notions of fairness are nonconsequentialist . . . . For example, if A wrongfully injured B, then B should be

---

189. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 40–41 n.53.

compensated by A; if A's action toward B would break a promise, then it is impermissible; if the true nature of the crime was  $X$ , then the punishment should be  $P(X)$ . In particular, whether A should compensate B, A may break a promise, or the punishment should be  $P(X)$  does not depend exclusively (or at all) on an assessment of the consequences of doing these things, such as the deterrence of undesirable behavior.<sup>190</sup>

Adding assumption (a) to get the result that a Fairness theorist is logically required (at least sometimes) to adopt a policy that would make everyone worse off *severely* restricts the scope of that result for KS. Only consequentialist Fairness theorists could be, by the logic of their own positions, required to conclude that they are sometimes morally required to make everyone worse off. And, as KS themselves point out, most Fairness theorists are not consequentialists.<sup>191</sup>

KS must also add assumption (b). If the agent is never faced with a choice between bringing about V or bringing about W', then it is hard to see how her judgment that V (where everyone is worse off) is a better situation than W' (where everyone is better off) would ever logically commit her to make any particular policy choice that she actually faced. And KS's need to assume (b) should not be underestimated, especially when one considers that the only way that an agent could ever actually be faced with a choice between bringing about V or bringing about W' is if she were actually able to distribute a good that had the features that (A1) and (A2) contemplate. This is implausible in the extreme because such a good would need, at least, to be infinitely divisible; such that the possession of any additional nonzero amount of it (no matter how microscopic) made any person better off; and capable of being distributed by the agent to everyone in the right amount to guarantee that W' is "constructed" properly.

#### *B. KS's Pareto Argument Provides No Reason to Believe (2)*

Does KS's proof provide any reason to think that welfarism is true? As explained above, KS's Pareto proof *does* demonstrate that any Fairness SWF for which (O), (A1), and (A2) are true violates (P), the weak Pareto principle. If we let Z stand for any such Fairness SWF, then we can represent the main logical components of KS's Pareto argument in terms of three premises establishing a conclusion, (C):

---

190. Kaplow & Shavell, *Notes on the Pareto Principle*, *supra* note 3, at 334.

191. *Id.*

(O) & (A1) & (A2), therefore;

(C): Z violates (P) for at least one pair-wise comparative evaluation of logically possible states.

Recall that welfarism is the view that the only features of a state that determine the state's intrinsic value are, collectively, the state's utility information.

Does welfarism follow deductively from (C)? Not in any obvious sense. So the idea that the Pareto argument in any clear sense demonstrates welfarism is ruled out immediately. But perhaps KS's proof could be enriched somehow to demonstrate not only (C) but welfarism as well. And if we can enrich KS's proof to create an argument for welfarism, then the enriched version of KS's proof can provide a reason to think that welfarism is true, especially if our enrichment requires only the addition of premises that are plausible. Perhaps the place to begin is to ask why it is such a bad thing for the Fairness theorist that there is a pair of logically possible states for which Z violates (P). What, exactly, is so horrible about violating the weak Pareto principle? Why is a SWF like Z problematic if, for at least one pair of logically possible states V and W', the SWF says that V is better than W' even though everyone is worse off in V than they are in W'?

By adding the assumption that (P) is *true* to KS's proof we might be able to articulate a reason why Z is problematic. KS's demonstration does *not* assume that (P) is true. The demonstration's only premises are (O), (A1), and (A2), and its conclusion is that any suitably "continuous" Z violates (P). If we add the premise that (P) is *true*, though, we can demonstrate at least the following new result: Z yields at least one incorrect comparison. For Z would say that V is better than W', when *in fact* V is worse than W'. We can help ourselves to the idea that V is in fact worse than W' once we assume (P), that any (logically possible) situation in which everyone is worse off must be a worse situation than one in which everyone is better off. And the weak Pareto principle seems pretty plausible, so we may be on our way. By adding the premise that (P) is true, we can conclude that Z yields at least one incorrect pair-wise comparison, even though at least one of the states that Z erroneously compares (namely W') is almost certainly infeasible (because for W' to be actual, there really would need to exist a good of the kind that KS imagine in their "mathematical" construction of W' and we really would need to be able to distribute it to everyone in just the right amount). But does welfarism follow from the conclusion that Z yields at least one incorrect comparison of two logically possible

states? The conclusion that Z yields at least one such incorrect comparison does not, in itself, seem to further KS's chances of showing that the only features of any state that are relevant to that state's value are, collectively, the state's utility information. What, if anything, is really doing the work here?

We must keep in mind why we can now say that Z is flawed. Z is flawed because we are now assuming that (P) is true. That is, we are now assuming that, if everyone's utility (well-being) is higher by some amount (not necessarily the same per person) in state W' than in state V, then W' *must be* better than V. This may seem a plausible assumption because it may seem odd to say that it would ever be *possible* for a situation in which everyone is worse off to be better than a situation in which everyone is better off. But if someone really was on the fence as to whether a situation's utility information was the only thing that *could ever* affect that situation's value, would that person, upon reflection, think that (P) is obviously true? *If* welfarism is true, *then* the weak Pareto principle almost certainly is. But if one is not antecedently committed to welfarism's being true, then what compelling reason is left for endorsing the weak Pareto principle?

To clarify this last point, if something other than individual utilities *could* affect the value of a situation, then why be so sure that any *logically possible* situation in which everyone's individual utility is higher *must always* be a better situation? (I am *not* arguing here that it follows from welfarism's being false that (P) is also false. Welfarism is a stronger claim than (P). Instead, I am arguing that it is hard to understand why someone would be tempted to endorse (P) if something *other* than individual utility ever even *could* affect the comparative values of situations.) The weak Pareto principle, after all, is a fairly strong claim. According to (P), it is *logically impossible* for a situation in which everyone is worse off to be better than one in which everyone is better off. If it is *logically possible* for something other than individual utilities to affect the comparative values of states, then what *plausible* reason is left to think that it is *logically impossible* for a state in which all individual utilities are higher to be less valuable than one in which they are all lower?<sup>192</sup>

---

192. Although I just argued that KS's Pareto argument provides no reason to think that welfarism is correct, I think that KS would concede this and point out that their Pareto argument was not designed to do so.

C. *KS's Pareto Argument Provides No Reason to Believe (3)*

KS themselves (correctly) acknowledge that the Pareto argument does not provide *any* independent support for (3):

We . . . note that much of our more formal argument has an even broader application. Namely, if an analyst thought that a concept of well-being that was qualitatively different from the welfare economic one . . . was normatively compelling, important parts of our analysis would still hold. In particular, there would be no change in the logic of our argument that giving any weight to a notion of fairness that is independent of well-being always raises the possibility that everyone would be made worse off; everyone being made worse off would be interpreted by reference to the analyst's own conception of individuals' well-being.<sup>193</sup>

As KS clarify, their Pareto argument neither presupposes nor gives any reason to endorse welfare economics' preference-satisfaction view about individual well-being.<sup>194</sup>

D. *KS's Pareto Argument Provides No Reason to Believe (4)*

It is important to continue to keep in mind that for KS's project to succeed, KS must provide *at least some* support for each of the following:

- (1) Welfare-economic consequentialism is true.
- (2) Welfarism is true.
- (3) The "degree-of-preference-satisfaction" account of individual well-being is true.
- (4) Legal and policy analysts *should* engage exclusively in welfare-economic critical-level thinking when generating and justifying normative (i.e., prescriptive or proscriptive) legal and policy scholarship.

And supporting (1)–(3) is crucial to supporting (4). That is, if an argument provides no reason to believe any one of (1), (2), or (3), then it provides no reason to believe (4).

---

193. KAPLOW & SHAVELL, FAIRNESS VERSUS WELFARE, *supra* note 3, at 23 n.14.

194. *Id.*

To see why, it is helpful to recall the distinction between critical-level and intuitive-level moral thinking.<sup>195</sup> As previously mentioned, intuitive-level thinking is non-inferential and often emotionally-infused.<sup>196</sup> When we form moral judgments intuitively, the judgment is not the result of a process of argumentation or analysis.<sup>197</sup> An analogy to perception is helpful here. I judge (believe) that as I type this sentence there is a keyboard in front of me. My keyboard judgment is not the result of inferential reasoning. It is instead what one might call a “perceptual belief.” I judge that there is a keyboard because I perceive it—I see, feel, and hear that it is there. Intuitive-level moral judgments are often made in (very roughly) the same way—we judge that such-and-such is morally wrong because we intuit that it is (*not* infer that it is).<sup>198</sup> And just as I might doubt the veracity of my perceptual beliefs if I have reason to doubt my senses, I might doubt the veracity of my intuitive moral judgment that, say, Negligent Nancy should compensate Adam for the harm her negligence causes if I have reason to doubt the reliability of the part of my brain that generates my moral intuitions.<sup>199</sup>

In contrast, proper critical-level (explicitly consequentialist) moral thinking avoids intuitive, emotional entanglements.<sup>200</sup> As explained, the critical-level thinker calculates—she identifies every feasible option, determines the consequences of each, and then bases her decision solely on which consequences are the best, for only that decision is morally permissible.<sup>201</sup> R.M. Hare offers a strikingly stark characterization of this sort of thinking:

[N]o moral intuitions of substance can be appealed to. It proceeds in accordance with canons established by philosophical logic and thus based on linguistic intuitions only. To introduce

---

195. *See supra* Part II.A.

196. *See supra* Part II.A.

197. *See supra* Part II.A.

198. *See supra* Part II.A.

199. Note that this Article assumes that cognitivism (as to moral judgments) is true and that moral realism is true. In other words, it is assumed that moral judgments can be correct or incorrect and that there are non-subjective facts of the matter as to whether, say, a given action is morally required. Just as a very myopic person might erroneously judge that there are no trees within 100 feet of her current position when in fact there are, a person with a defective moral sense might erroneously judge that a certain option is morally permissible when in fact it is not. These are enormous meta-ethical assumptions, but defending them would require extensive argumentation beyond this Article’s scope.

200. *See supra* Part II.A.

201. *See supra* Part II.A.

substantial moral intuitions at the critical level would be to incorporate in critical thinking the very same weakness which it was designed to remedy.<sup>202</sup>

Put more crisply, “[c]ritical thinking consists in making a choice under the constraints imposed by the logical properties of the moral concepts and by the non-moral facts, and by nothing else.”<sup>203</sup>

This may seem an unacceptably skeletal description of a form of *moral* thinking. One might ask how any form of moral reasoning could proceed devoid of any substantive normative or evaluative commitments. It is extremely implausible that Hare means to exclude substantive views about what makes one situation better or worse than another from playing any role in critical-level moral thinking. Imagine a form of consequentialism that is paired with no commitment whatsoever as to what makes one situation better than another. For example, start with a sophisticated welfarist who also endorses consequentialism, but then take away the welfarism part—the part where she thinks that individual utility is the only source of ethical value that a situation has. Once that substantive part is stripped away, the “denatured” consequentialist would have nothing to say about what makes a choice morally permissible other than “an available option that, as a consequence, maximizes something.” Without the substantive commitment, welfare economics loses the idea that increasing individual utility counts in favor of a particular alternative.

It seems most plausible to interpret Hare as articulating an aspect of explicitly consequentialist thinking that is independent of any *particular* view about what makes a situation better or worse. That is, to engage fully in critical-level moral thinking, one must have *some* substantive commitment about what makes one outcome better than another. But there is no respect in which any morally substantive commitment enters into critical-level thinking that is independent of what the reasoner’s view about the source of a situation’s value happens to be. For critical-level moral reasoning, everything *other than* considerations concerning the value of a situation boils down to matters of fact about logical/semantic properties of propositions and about “what causes what.”

This characterization of critical-level thinking implies that there are two factors that determine whether a given agent *should* engage in

---

202. HARE, *supra* note 22, at 40.

203. *Id.*

critical-level thinking when addressing a normative issue: (a) how good the agent is at executing consequentialist analysis (identifying what causes what, identifying all of the feasible options, drawing valid logical inferences, etc.) and (b) whether the agent's underlying theory about what makes a situation ethically valuable is true.<sup>204</sup> It is for this reason that if an argument provides no reason to believe any one of (1), (2), or (3), then it provides no reason to believe (4). And (1)–(3) *are* welfare economics. If there is no reason to think that consequentialism is true, no reason to think that welfarism is true, and no reason to think that the preference-satisfaction view of individual well-being is true, then there is no reason to think that KS's normative claim about how legal analysts should proceed vis-à-vis normative issues is true.

As explained, to address an agent's ability to engage in explicit consequentialist reasoning, Hare often uses the idea of an archangel—a superhuman being not subject to limits on human rationality and other human weaknesses (e.g., the tendency to be self-serving).<sup>205</sup> Archangels never make logical errors, and they can infallibly predict the consequences of any feasible alternative in any choice situation.<sup>206</sup> If there is no reason to think that welfare economics (i.e., (1)–(3)) is true, then there is no reason why *even an archangel* should engage in welfare-economic critical-level thinking. Perfect consequentialist reasoning that transparently tracks the evaluative criteria of a potentially flawed moral theory will yield moral judgments that may well be false.

For these reasons, KS's Pareto argument provides no reason to think that *any one* of (1), (2), (3), or (4) is true. And (1)–(4) constitute KS's project. *If* an argument provides no independent reason to accept any part of a project, *then* that argument cannot properly be considered an argument in favor of that project. Therefore, KS's Pareto argument should not be considered among the arguments that favor KS's project, even a little.

---

204. See *supra* Part II.A.

205. See *supra* Part II.A; see also HARE, *supra* note 22, at 44–45.

206. See *supra* Part II.A; see also HARE, *supra* note 22, at 44–45.