

The Pigou-Dalton Principle and the Structure of Distributive Justice

Matthew D. Adler

Richard A. Horvitz Professor of Law and Professor of Economics, Philosophy and Public Policy
Duke University. adler@law.duke.edu

Working paper, May 2013

The Pigou-Dalton (PD) principle,¹ applied to some type of good, recommends a non-leaky, non-rank-switching transfer from someone with more of the good to someone with less, as long as no one else's holdings are changed—non-leaky in the sense that the one who starts out with less of the good gains by exactly as much as the other loses; non-rank-switching in the sense that the one who starts out with less does not end up with more than the other. Here is a formal statement of this principle: Let $g = (g_1, g_2, \dots, g_N)$ be a list of the holdings of good G among the population of N individuals, with g_i a number quantifying the holdings of any given individual i , and $g^* = (g_1^*, g_2^*, \dots, g_N^*)$ another such list. If there exist two individuals h and l such that $g_h > g_l$, $\Delta > 0$, $g_h^* = g_h - \Delta \geq g_l^* = g_l + \Delta$, with $g_i = g_i^*$ for every other individual i in the population, then g^* is an improvement over g .

In this Article, I defend the PD principle as a principle of *distributive justice*, with the relevant good *responsibility-adjusted well-being*. Roughly speaking, the principle I will defend says: if one person is at a higher level of well-being than a second, and the worse-off one is not responsible for being worse off, then distributive justice recommends a non-leaky, non-rank-switching transfer of well-being from the first to the second, if no one else's well-being changes.

The PD principle has been little discussed in the philosophical literature. This is surprising. The principle is the core of the economic literature on measuring inequality. Here, the good is income, and economists typically take as axiomatic that an inequality metric should register a Pigou-Dalton transfer in income as reducing the degree of income inequality.²

Of course, there are many economic concepts that don't surface in philosophy, and vice versa. But it's not as if philosophers have no interest in equality. Nor is it that the economists have endorsed a principle which, after seriously philosophical reflection, seems unattractive. Just the opposite. The PD principle, suitably framed, is a very plausible principle of distributive justice—or so I shall argue.

¹ Originally suggested by Arthur Pigou, *Wealth and Welfare* 24 (New York: Macmillan, 1912), and Hugh Dalton, "The Measurement of the Inequality of Incomes," *Economics Journal* 30 (1920): 348-61.

² See, e.g., Satya Chakravarty, *Inequality, Polarization and Poverty* (New York: Springer, 2009), ch. 1; Frank Cowell, *Measuring Inequality*, 3d ed. (Oxford: Oxford University Press, 2011); Bhaskar Dutta, "Inequality, Poverty and Welfare," in *Handbook of Social Choice and Welfare*, vol. 1, ed. Kenneth Arrow et al. (Amsterdam: Elsevier, 2002), 597-633.

It might be objected that philosophers *have* widely discussed the PD principle, just with a different name: “prioritarianism.” Prioritarians say that benefits to worse-off individuals have greater moral weight. This is just to endorse the PD principle, applied to well-being, as a principle of morality.

While the PD principle (in this form) certainly *is* a defining commitment of prioritarianism, prioritarians also typically seem to embrace an additional principle—call it “separability”—which is logically distinct from the PD principle, neither implying nor implied by it.³ In his seminal presentation, Derek Parfit characterizes prioritarianism as having the feature that the moral value of someone’s benefit does not depend upon her position relative to others.⁴ The best axiomatic rendering of this feature is separability. Although prioritarians are often fuzzy about the difference between the PD principle and separability, the two principles *are* distinct.

Further, separability is more contestable than the PD principle. While I do find separability to be an attractive principle of distributive justice, the case for endorsing the PD principle is yet stronger than the case for endorsing the combination. This Article seeks both to make progress in our understanding of the structure of distributive justice, and to clarify the content of prioritarianism, by arguing for the PD principle while bracketing the question of separability.⁵

I defend the PD principle with reference to a particular *justificatory framework* for distributive justice: a constellation of concepts and propositions that helps to identify correct principles of justice, and to explain their correctness.⁶ Call this the “benefit-claim” framework—a framework adumbrated in the work of Thomas Nagel on egalitarianism. Each individual has a claim in favor of a distribution of goods that makes her better off; one distribution is more just than a second iff it more fairly accommodates everyone’s potentially conflicting claims.

It would be nice if *every* plausible justificatory framework supported the PD principle. Unfortunately, I do not believe this to be true. Two other popular frameworks are the veil-of-ignorance framework of John Rawls and John Harsanyi; and the “complaint” framework developed by Larry Temkin. Neither clearly supports the PD principle.

³ See below, Part IV.

⁴ Derek Parfit, “Equality or Priority?” in *The Ideal of Equality*, ed. Matthew Clayton and Andrew Williams (Houndmills: Palgrave, 2000), 81-125.

⁵ In chapter 5 of my book, *Well-Being and Fair Distribution* (Oxford: Oxford University Press, 2012), I argue for both the PD principle and separability, and thus prioritarianism. This Article presents a substantially refined version of the argument in the book for the PD principle—seeking here to show why the case for PD is particularly powerful. My own views continue to be prioritarian; but it is important as a deliberative matter to differentiate between PD and separability, and to see how the case for the PD principle flows from the benefit-claim framework in an especially direct way.

⁶ The arguments in this Article are neutral on metaethical questions, and “correct” is just a shorthand, which can be given a suitably noncognitivist interpretation by those who deny or deflate moral facts and truths.

The strategy here, therefore, is indirect. First, I present a powerful argument *from* the benefit-claim framework *to* the PD principle, suitably specified. Second, I seek to undercut the two competing frameworks just mentioned, arguing that each (despite its popularity) is a problematic way to organize our thinking about distributive justice.

The topic of distributive justice has many aspects, including these: What *is* the good whose distribution is governed by principles of justice?⁷ Do these principles govern distributions between individuals who are members of separate societies, or only within societies?⁸ Does justice apply only at the level of society's "basic structure," or to individuals' day-to-day choices?⁹

All these questions have been intensively mooted in scholarship about justice over the last several decades. The question addressed in this Article has been less fully discussed: What are the criteria for ranking distributions in light of justice? A perfectly equal distribution is more just than an unequal distribution of the same sum total of goods; but what else can be said?¹⁰ The PD principle, I suggest, is one important part of the answer to this question.

I. *The PD Principle: A Statement*

I take it as uncontroversial that the subject of distributive justice is, at least, *distributions*. Allocation of goods (of some sort) within certain groups of individuals are assessable as more or less just.

More specifically, my discussion of distributive justice will employ the following set-up. Let a "justice population" be some group of individuals, suitably related so that the allocation of goods among them is assessable as more or less just. Let a "good" be any individual attribute (monadic or relational) that contributes to individual well-being. Let a "distribution" among some justice population be a full specification of the goods held by the members of the population, sufficient to determine how well off each such individual is; and a full specification

⁷ See, e.g., Richard J. Arneson, "Welfare Should be the Currency of Justice," *Canadian Journal of Philosophy* 30 (2000): 497-524.

⁸ See, e.g., Martin O'Neill, "What Should Egalitarians Believe?" *Philosophy & Public Affairs* 36 (2008): 119-56, esp. 134-39.

⁹ See, e.g., A.J. Julius, "Basic Structure and the Value of Equality," *Philosophy & Public Affairs* 31 (2003): 321-55.

¹⁰ To be sure, a related question *has* been intensively discussed, namely; what are the criteria for ranking outcomes and choices in light of *morality*? But those discussions have gone well beyond the topic of justice, since many of the philosophers engaged therein believe that considerations other than justice are also *moral* considerations—for example, compassion, virtue, or benevolence. See, e.g., Roger Crisp, "Equality, Priority, and Compassion," *Ethics* 113 (2003): 745-63; Larry Temkin, "Equality, Priority, and the Levelling Down Objection," in *The Ideal of Equality*, ed. Matthew Clayton and Andrew Williams (Houndmills: Palgrave, 2000), 126-61; Shelly Kagan, *The Geometry of Desert* (Oxford: Oxford University Press, 2012). Prioritarianism, in particular, was chiefly presented by Parfit in "Equality or Priority" as a complete account of moral value—not a view about justice in particular—and that is how it is usually discussed.

of “responsibility facts” about each member, sufficient to determine how prudently she has behaved. I assume that prudence levels can be given a complete ranking.¹¹

Let a “justice ranking,” relative to some justice population, be a quasiordering (a transitive, possibly incomplete ranking) of distributions among that population. The total set of possible distributions among a given justice population can be partitioned into a series of subsets, within which each individual’s prudence level is fixed. (That is, for each distribution within such a subset, individual 1 is at prudence level L_1 , individual 2 at level L_2 , etc.). I focus throughout this Article on the simpler problem of how a justice ranking should order distributions within any such subset; the proviso that prudence levels are fixed will be implicit in what follows.¹²

The version of the PD principle about to be stated is formulated for this set-up, and for the case of fixed prudence. Many other versions can be formulated; but since those variations are not discussed here, I can without confusion refer to this one as the “PD principle.” A “principle of distributive justice,” such as the PD principle, helps to specify the justice ranking—by articulating sufficient conditions for two distributions to be equally just, for one to be more just than another, or for the two to be non-comparable with respect to justice.

The PD Principle

Assume that there are two distributions d and d^* , and two members of the population (call them “High” and “Low”), which satisfy the following three conditions.

- (1) The Level Condition. In d , High is better off than Low. In d^* , High is at least as well off as Low.
- (2) The Delta Condition. High is better off in d than d^* , while Low is better off in d^* than d , and the difference in High’s well-being between the two distributions is exactly equal to the difference in Low’s well-being.
- (3) The Responsibility Condition. High and Low behave equally prudently in d and d^* , and this prudence level is sufficiently high.

¹¹ Consider the set \mathbf{P} of possible pairings of individuals and distributions, i.e., the product set of the justice population and the set of distributions. Assume that, for any two such pairings, either the first individual in the first distribution is more prudent than the second individual in the second distribution, or vice versa, or they are equally prudent. Then \mathbf{P} can be partitioned into a series of equivalence classes with respect to prudence. Each such class defines a prudence level, and the levels are completely ordered.

The simplifying assumption of a complete ranking of prudence levels, in turn, allows us to partition the set of *distributions* into a series of mutually exclusive subsets, within each of which any given individual is at a fixed prudence level.

¹² Difficult puzzles arise if we ask whether a change in how prudently someone behaves would itself be more or less just. This Article focuses on the simpler question: to what degree would various arrangements of goods among a group of individuals be just, given that the individuals have exhibited or will exhibit a particular pattern of prudence?

Moreover, the following condition is true of every other individual in the justice population. (As a shorthand, I will say that someone is “unaffected” by two distributions if she is equally well off in each, and “affected” if this is not the case.)

(4) The Others Unaffected Condition. Everyone else in the population is equally well off in d^* as in d .

Then: Distribution d^* is more just than d .

When are individuals “suitably related” to be governed by distributive justice? What does it take for a group of individuals to be a “justice population?” The set-up just presented, and the defense of the PD principle about to be put forward, are agnostic on this issue.

This set-up and defense are also agnostic about the kinds of choices to which principles of justice apply: individuals’ day-to-day choices, choices with respect to a society’s “basic structure,” etc. *Some* choice situations, at least, are governed by distributive justice; and the justice ranking of distributions, in turn, will help fix what justice does require in such situations.

The reader will note that a distribution is a full specification of the characteristics of a justice population, sufficient to determine how well off they are and how responsibly (prudently) they’ve behaved. A distribution is not quite a possible world. There will be multiple worlds compatible with a given distribution: a set of worlds any one of which could possibly obtain, if the distribution were to. However, those worlds will differ only with respect to facts that are irrelevant to both well-being and prudence. Any given individual will be equally well off in all of the worlds compatible with a distribution, and will have behaved equally prudently.

My set-up employs “distributions” with this full-specification property so as to avoid difficult questions regarding the application of the PD principle under conditions of risk and uncertainty.¹³ However, given the bounded capacities of the human mind, such “distributions” are cognitively intractable objects for human agents. A human mind is not able to store an explicit representation of a distribution.

The topic addressed here therefore concerns *criteria of rightness* rather than *decision procedures for humans*. In particular, I am asking: what are the criteria of rightness, in light of distributive justice, for one kind of object to which criteria of justice are applicable, namely fully specified distributions? Once we have seen (as I hope to demonstrate) that the PD principle figures among the criteria of justice for fully specified distributions, we can then engage related topics: What are the criteria of rightness for *choices* governed by distributive justice, or for incompletely specified distributions? Insofar as *these* objects correspond to probability distributions across fully specified distributions, does the PD principle govern *their* ranking in an “ex post” or “ex ante” manner? These question, in turn, are different from this one: what role

¹³ See Adler, *Well-Being and Fair Distribution*, ch. 7.

does the PD principle play in the decision procedures that cognitively bounded agents ought to employ in making those choices that are subject to the requirements of justice?

These latter topics, although vitally important, are placed to one side. It will be progress enough to show that the PD principle states a sufficient condition for one fully specified distribution (I henceforth drop the phrase “fully specified,” which is implied) to be more just than a second.

The PD principle, if indeed a true principle of justice, may well help to justify additional principles—in particular, additional principles governing transfers. For example, the PD principle together with a plausible principle of “Anonymity” entails an expanded PD principle—endorsing a non-leaky transfer that may be rank-switching but is gap-diminishing.¹⁴ A different kind of expanded PD principle broadens the Responsibility Condition so that Low is either equally prudent as High or more prudent. The PD principle may also help to justify transfer principles concerning *leaky* transfers—stating conditions under which improving a worse-off person’s welfare by *less* than is lost by a better-off person yields a more just distribution.

Thus the PD principle is certainly *not* a necessary condition for one distribution to be more just than a second. Rather it is a *sufficient* condition—or so I argue here.

While my defense of that principle is agnostic about various questions that I have just described, it is *not* agnostic about the “currency” for distributive justice. I take that currency to be responsibility-adjusted well-being¹⁵—not primary goods, resources, midfare, or capabilities, to name the obvious competitors. “Goods” are defined as attributes constitutive of well-being. The Level Condition and Delta Condition are expressed in terms of the well-being levels of High and Low, and changes in their well-being, and not in terms of their levels of primary goods, resources, midfare, capabilities, etc. Moreover, the “currency” for the set-up and principle is obviously *responsibility-adjusted* welfare, not welfare simpliciter. The PD principle does not make the combination of the Level, Delta, and Others Unaffected Conditions sufficient for d^* to be more just than d . Those conditions are only jointly sufficient for that consequence *together with* the Responsibility Condition.

The discussion here will not reshuffle the “equality of what” debate, but instead will presuppose that “responsibility-adjusted well-being” wins that debate. In showing how principles of distributive justice flow from the pattern of individuals’ benefit claims, and how the strength of such claims can readily be grounded in facts about individuals’ welfare and

¹⁴ The expanded PD principle uses a different version of the Level Condition: High is better off than Low in d , and may be worse off in d^* , but the difference (“gap”) between the two individuals’ well-being in d^* is less than the difference in d . “Anonymity” says that if distributions d^+ and d^{++} are identical in the pattern of distribution of well-being (the worst-off person in d^+ , second-worst off, etc., is equally well off as her counterpart in d^{++}), and in the association of prudence levels with well-being levels—although differing perhaps in the *names* of the individuals at particular well-being and prudence levels—then d^+ and d^{++} are equally just.

¹⁵ See Arneson, “Welfare should be the Currency of Justice.”

responsibility, the discussion buttresses existing arguments for that “currency”; but no attempt will be made to recapitulate those arguments.

But what of readers who prefer primary goods, capabilities, midfare, resources, etc.? Is there a version of the PD principle that *they* should embrace? This is an important question—the “equality of what?” debate remains a *debate*—but not one I will attempt to address. One worry about the PD principle as a generic, currency-independent principle of distributive justice has to do with the Pareto principle. My defense of the PD principle (the version just presented) will rely upon the Pareto principle. But the PD principle, in *some* non-welfare currencies, conflicts with that principle.¹⁶

II. *Why Endorse the PD Principle? The Benefit-Claim Framework*

Why believe that the PD principle is true? Here, I draw upon Thomas Nagel’s critique of utilitarianism and defense of what he terms an “egalitarian priority system.” In his 1977 Tanner Lecture, he writes:

[An egalitarian priority system, unlike utilitarianism] does not combine all points of view by a majoritarian method. Instead, it establishes an order of priority among needs and gives preference to the most urgent, regardless of numbers. . . .

One problem in the development of this idea is the definition of the order of priority: whether a single, objective standard of urgency should be used in construing the claims of each person, or whether his interests should be ranked at his own estimation of their relative importance. In addition to the question of objectivity, there is a question of [temporal] scale . . .

But let me leave these questions aside. The essential feature of an egalitarian priority system is that it counts improvements to the welfare of the worse off as more urgent than improvements to the welfare of the better off. These other questions must be answered to decide who is worse off and who is better off, and how much, but what makes a system egalitarian is the priority it gives to the claims of those whose overall life prospects put them at the bottom, irrespective of numbers or of overall utility. Each individual with a more urgent claim has priority, in the simplest version of such a view, over each individual with a less urgent claim. The moral equality of egalitarianism consists in taking into account the interests of each person, subject to the same system of priorities of urgency, in determining what would be best overall.¹⁷

In *Equality and Partiality*, Nagel writes:

[I]mpartiality generates a greater interest in benefitting the worse off than in benefitting the better off—a kind of priority to the former over the latter. . . .

[Individualized impartial concern] does not rule out all ranking of alternatives involving different persons,

¹⁶ See Adler, *Well-Being and Fair Distribution*, 114-19; Marc Fleurbaey and Alain Trannoy, “The Impossibility of a Paretian Egalitarian,” *Social Choice and Welfare* 21 (2003): 243-63; Marc Fleurbaey, “Social Welfare, Priority to the Worst-Off and the Dimensions of Individual Well-Being,” in *Inequality and Economic Integration*, ed. Francesco Farina and Ernesto Savaglio (London: Routledge, 2006), 225-68.

¹⁷ “Equality,” in *Mortal Questions* (Cambridge: Cambridge University Press, 1979), 106-227, 117-18.

nor does it mean that benefitting more people is not in itself preferable to benefitting fewer. But it does introduce a significant element of non-aggregative, pairwise comparison between the persons affected by any choice or policy. . . . The claims on our impartial concern of an individual who is badly off present themselves as having some priority over the claims of *each* individual who is better off: as being ahead in the queue, so to speak.¹⁸

Several key ideas animate these passages. First, the egalitarian ranking of outcomes is determined by the totality of individuals' "claims" (a word Nagel uses repeatedly). Second, what each individual can claim is an "improvement" in her well-being—that she be "benefitted." Third, individuals who are worse off have stronger claims. Nagel also suggests, fourth, that alternatives are to be compared in a "non-aggregative" manner, by pairwise comparison between claims even when more than two individuals have claims; but this further suggestion will not figure in my defense of the PD principle.

Drawing to begin upon the first two ideas, I suggest the following framework—the "benefit-claim" framework—for grounding principles of distributive justice. A claim is a relation between an individual and two distributions. Any claim has a *valence*: it is either a zero claim, or a claim in favor of one distribution over the other.¹⁹ Every non-zero claim also has a *strength*. The ranking of two distributions is determined by some rule (aggregative or not) for comparing the number and strength of non-zero claims in favor of the first distribution, with the number and strength of non-zero claims in favor of the second.

Many different rules for assigning claim valence and strength can be imagined. I assume, to start, only that a well-being difference is a necessary if perhaps not sufficient condition for a non-zero claim. If Sue is equally well off in d and d^* , how could it be unfair to her if one distribution rather than the other obtained?²⁰

The argument from the benefit-claim framework (with this proviso about valence), to the PD principle, is straightforward. To begin, High surely has a non-zero claim in favor of d over d^* . After all, High is better off in d than d^* . Now, if High's behavior in one or both distributions had been badly imprudent, we might assign him a zero claim despite this well-being difference. We might, in that case, say that distributive justice is indifferent to welfare effects on

¹⁸ *Equality and Partiality* (New York: Oxford University Press, 1995), 66-68.

¹⁹ My presentation, for simplicity, ignores incomparability in the well-being ranking—any individual is either better off in one of two distributions, or equally well off. The benefit-claim framework could easily be refined to allow for well-being incomparability. See Adler, *Well-Being and Fair Distribution*, ch. 5. If someone is neither better off with one distribution than a second, nor equally well off, she can be assigned a claim with an indeterminate valence. This version of the benefit-claim framework still supports the PD principle—which remains as stated above. Note that the PD principle itself identifies a circumstance where no one (neither High and Low, nor the rest of the population) is incomparably well off in the two distributions, and thus no one has a claim with an indeterminate valence.

²⁰ To be sure, *non-welfarists* might see unfairness in Sue's having one bundle of "goods" (understood in some non-welfarist sense) rather than a second, even if the bundles are equally good for her well-being; but the aim here as already explained is to trace out an argumentative path to the PD principle from an initial starting point of responsibility-adjusted welfarism—not to justify the starting point.

someone so irresponsible. But one aspect of the Responsibility Condition is that High is *sufficiently* prudent in both d and d^* . At a minimum, if High is perfectly prudent in both distributions, then it would be problematic for an account of distributive justice to give no role to High's interests in the determination of whether d or d^* is more just. The level of "sufficient prudence" might be at the level of perfect prudence, or lower, all the way down to maximal imprudence; nothing here hinges upon where it is set.

Parallel reasoning shows that Low has a non-zero claim in favor of d^* over d .

Moreover, because everyone else is equally well off in d and d^* (the Others Unaffected condition), everyone else has zero claims. Thus we have a two-person case, in which one person (High) has a non-zero claim in favor of d , and another (Low) a non-zero claim in favor of d^* .

Who has the stronger claim? Nagel suggests, I have noted, that worse-off individuals have comparatively stronger claims. ("Improvements to the welfare of the worse off [are] more urgent than improvements to the welfare of the better off."²¹) Very plausibly, this is true in a weak, pro tanto form. If no other factors determinative of claim strength count the other way, then: if Sam would be better off than Sheila, regardless of which distribution obtains, then Sheila has a stronger claim between the distributions than Sam. Extending this slightly, it seems very plausible to endorse the following rule for assigning claim strength.

The Pro Tanto Well-Being-Level Rule for Claim Strength: If one individual is better off than a second in distribution d^+ , and either better off than or equally well off as the second in distribution d^{++} , then—if no other considerations relevant to claim strength count the other way—the first individual has a weaker claim between d^+ and d^{++} than the second individual.

Various additional factors may determine the strength of an individual's claim between two distributions, potentially including: her well-being difference between the two; how prudently she and others behave in each; the overall pattern of well-being in each. The pro tanto rule just stated merely says that, where no other such factors are overriding, the first individual has a weaker claim between the distributions than the second.

But consider the case of two distributions d and d^* satisfying the conditions of the PD principle. Because Low is worse off than High in d , and worse off or equally well off in d^* , her pro tanto claim between the two is stronger than High's. Moreover, those conditions ensure that no other factors operate to yield a stronger all-things-considered claim for High than Low. By virtue of the Delta Condition, High's well-being difference between d and d^* is exactly equal to Low's well-being difference between d^* and d . In a different case, if High had sufficiently *more* to gain than Low to lose, High might plausibly argue that the distribution which makes him better off is on balance the fairer distribution. But he can hardly make such an argument here.

²¹ "Equality," 117-18.

The Responsibility Condition states that High and Low are not merely sufficiently, but also equally, prudent. Arguably, as between two individuals both of whom are affected by two distributions—and both of whom are sufficiently prudent to have non-zero claims—considerations of *comparative fault* should influence the comparative strength of their claims. For example, imagine that normal prudence is sufficient, and that Normal behaves at that level, while Perfect is perfectly prudent. In one distribution, Perfect is worse off than Normal; in a second, they swap places; no one else is affected. In such a case, even though the well-being levels are symmetrical and the differences equal, wouldn't Perfect have stronger grounds for complaint if the first distribution were to obtain, then Normal would if the second were?

However, in the case at hand, even if Low falls short of perfect prudence, High does so to an equal extent. Thus, considerations of comparative fault do not vitiate the pro tanto case for a stronger claim that Low has in virtue of being at a lower well-being level than High.

Finally, it is important to note that the argument presented here for the PD principle does *not* deny the potential relevance, for purposes of determining someone's claim strength, of the overall pattern of well-being levels in each of the two distributions. (We will return to this observation in the discussion, in Part IV, of separability.) But because High is better off than Low in d , it follows of course that High's *rank* in the distribution of well-being in d is higher than Low's rank in d : the number of individuals who are worse off than High in d is greater (by at least one) than the number of individuals who are worse off than Low in d . And High's well-being rank in d^* is either greater than or equal to Low's well-being rank in d^* . While a pro tanto rule giving a *stronger* claim to a higher-ranked person is possible, such a rule would be perverse. Either someone's position in the overall pattern of well-being makes no difference to her claim strength; or the fact that she has a higher position should tend to make her claim weaker.

In line with a standard analysis, the Responsibility Condition points to facts about High's and Low's prudence.²² Some have suggested that individuals can be differentially prudent but equally responsible for purposes of distributive justice, or vice versa. For example, Todd focuses like a laser beam on his own interests, while Teresa works tirelessly to advance distributive justice itself, sacrificing her own welfare. Slack is as heedless of his own welfare as Teresa, but for no good reason. All have ended up badly off; we can choose to benefit one. Wouldn't it be at least as fair to benefit Teresa as Todd, and fairer than to benefit Slack?

For the reader worried by this sort of example, it should be noted that the Responsibility Condition can be rendered more generic. High and Low are at an equal level of "effort," where "effort" means an individual's rationality in pursuit of some admissible mixture of goals,

²² Arneson, "Welfare Should be the Currency of Justice," 506-08; Kasper Lippert-Rasmussen, "Luck Egalitarianism: Faults and Collective Choice," *Economics and Philosophy* 27 (2011): 151-73, 169 (noting that "[t]he default view of faults [for luck egalitarians] is the view that the relevant evaluative standard is prudence, i.e., whether the agent conducts herself in a way that is relevantly worse than optimal from the point of view of the agent's self-interest").

including self-interest, moral goals, and perhaps others.²³ However, for concreteness I will stick with the Responsibility Condition as stated and analyze “effort” in terms of prudence.

III. *Why the Benefit-Claim Framework?*

Recall that a justificatory framework (for distributive justice) is a constellation of concepts and propositions that helps to identify correct principle of distributive justice. The benefit-claim approach is *one* plausible such framework, and it supports the PD principle. But other seemingly plausible justificatory frameworks do not—as the discussion to follow will show. While (as mentioned) economists studying inequality take the PD principle as foundational—and indeed I believe they are *correct* to do so—the defense of the PD principle is by no means trivial.

A full defense of that principle would seek to undercut every justificatory framework that fails to support it. That, of course, is well beyond the scope of this Article. Rather, I criticize two popular, competing frameworks: the veil-of-ignorance approach, and Temkin’s complaint framework.²⁴ The fact that a candidate framework fails to support intuitively compelling principles of distributive justice makes it difficult for the deliberator about justice to endorse the framework, in reflective equilibrium. But the veil-of-ignorance framework has difficulty supporting a principle of *Minimal Preference for Equality*, while Temkin’s framework has difficulty supporting the *Minimal Pareto Principle*. The benefit-claim framework supports both.

A. The Veil of Ignorance

While Rawls, of course, proposes principles of justice to be applied to the distribution of primary goods—not well-being—the welfarist can certainly appropriate the veil-of-ignorance framework by appeal to which Rawls justifies those principles. Define distributions as above, and consider the simple case where the subset of distributions being ranked is such that everyone’s (fixed) level of prudence is equal. In such a case, on the responsibility-sensitive welfarist construal of the veil of ignorance, one distribution is more just than a second iff each member of the justice population (or her guardian angel)²⁵—under an appropriate condition of ignorance concerning which attribute bundle she has, and choosing rationally—prefers the first.

²³ See Lippert-Rasmussen, “Luck-Egalitarianism,” 169-72. The formal literature on responsibility-sensitive egalitarianism adjusts for individual “effort,” without any particular substantive commitment about its content. See, e.g., John Roemer, *Equality of Opportunity* (Cambridge, MA: Harvard University Press, 1998).

²⁴ A third competitor is Tim Scanlon’s contractualism, whereby moral principles are those that individuals lack reason to reject. T.M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998). But Scanlon stresses that an individual’s reasons transcend her well-being, and thus his framework seems a poor basis for grounding principles of justice with a welfarist currency. Conversely, if the framework is adapted for this context by restricting a contractor’s reasons to her own well-being, it seems little different from the benefit-claim approach.

²⁵ Recall that we are developing criteria of rightness for *fully specified* distributions, and considering justificatory frameworks adapted to *this* problem. Because members of the justice population (if humans) will lack the cognitive resources to think in full detail about a distribution, the veil-of-ignorance framework should probably be formulated

Surely the justice ranking of distributions should satisfy a principle of *Minimal Preference for Equality*:

Minimal Preference for Equality: Assume that distributions d^+ and d^{++} are such that: (1) Total well-being (the sum total of individual well-being) is the same in both distributions. (2) In d^{++} , everyone is equally well off, while that is not true in d^+ . (3) Everyone is sufficiently prudent and equally prudent. Then d^{++} is more just than d^+ .

The veil-of-ignorance framework, under a very plausible plausible construal of an “appropriate condition of ignorance,” does not support this principle

Rawls argues that the veil should involve nonprobabilistic ignorance (“NI”).²⁶ If the veil in the context at hand is specified as an NI veil, then each member of the justice population compares the distributions without knowing which bundle of goods she will receive, *and* without assigning a probability (or even set of probabilities) to each given bundle. A maximin rule for rational choice under NI, combined with such a veil, will yield a maximin rule for ordering distributions—in turn satisfying Minimal Preference for Equality.

But Rawls’ case for NI is hardly compelling. John Harsanyi and, more recently, Derek Parfit have argued that an equiprobabilist construal of the veil of ignorance (“EI”) is at least as plausible as NI.²⁷ If there are N members of the population and thus each distribution allocates N bundles of goods to these persons, each individual—behind the EI veil—would view a given distribution as a $1/N$ chance of the first bundle, a $1/N$ chance of the second, etc. The best case for EI does *not* rely on a dogmatic Bayesianism, which says that rational individuals can always assign precise probabilities to the upshots of their choices, hence are never operating under NI. It allows that NI may be consistent, in some cases, with the dictates of rationality—but denies that NI provides the best interpretation of what “ignorance” is supposed to do as an element of a hypothetical choice procedure for arriving at principles of justice. The veil of ignorance is an *imagined* scenario, constructed by moral deliberators so as to enable clearer thought about the requirements of justice; the deliberator can stipulate that each bundle has precise probabilities, and can stipulate what they are. Moreover, ignorance is supposed to operationalize the separateness of persons—that each person’s interests have a distinct and equal role in determining a just allocation. Why isn’t such separateness best operationalized via EI, which tells the chooser behind the veil of ignorance to give equal—not unknown—weight to the possibility of each bundle?

in terms of a population of angelic (cognitively unbounded) advisers each caring about the interests of one member of the justice population. Since I am arguing *against* the veil-of-ignorance framework, I do not pursue this issue; the framework is problematic for quite a different reason, namely its failure to endorse Minimal Preference for Equality.

²⁶ John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Harvard University Press, 1999), 118-23, 130-53.

²⁷ John Harsanyi, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge: Cambridge University Press, 1986), ch. 4; “Morality and the Theory of Rational Behavior,” in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982); Derek Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), vol. 1, 350-51.

Does EI support *Minimal Preference for Equality*? Some have suggested that rational individuals, choosing with known probabilities, may be required—required as a matter of rationality—to be *risk-neutral* in well-being.²⁸ This means, first, that there is a well-being function $w(\cdot)$, representing the well-being levels of different attribute bundles, and well-being differences between them; and, second, that lotteries over bundles are ranked according to the mathematical expectation of $w(\cdot)$. A little less formally, this means indifference between a choice yielding a particular level of well-being with certainty, and any choice whose *expected* well-being value is that particular level. EI, together with this conception of rationality, clearly *violates* Minimal Preference for Equality. EI plus risk neutrality implies that each individual ranks distributions according to the *average* $w(\cdot)$ value of the component bundles—the average such value, since each bundle has an equal probability. But d^+ and d^{++} , as described by *Minimal Preference for Equality*, have the same average $w(\cdot)$ value. d^{++} is a perfect equalization of the same sum total of well-being that is attained, unequally, in d^+ .

To be sure, it is hardly compelling that rationality *requires* risk neutrality in well-being. Moreover, if that putative requirement is replaced with a different one—namely, that rationality requires risk *aversion* in well-being—the veil of ignorance plus EI *will* be consistent with Minimal Preference for Equality. Informally, risk aversion in well-being means caring more about the downside risk of low well-being levels, than the upside chance of high levels. Distribution d^+ (understood as an equiprobability distribution over bundles) offers both upside and downside risks, relative to d^{++} . And because the two distributions have the same expected well-being value, the downside risks outweigh the upside chances for any risk-averse chooser and incline her towards d^{++} .²⁹

But seeing rationality as requiring risk aversion in well-being is no more plausible than seeing it as requiring risk neutrality. The latter requirement errs, if it does, by precluding a diversity of rational risk attitudes. If it is rationally permissible to prefer chocolate or vanilla, why not to be risk-neutral, -averse, or risk prone in well-being? Yet shifting from a rational *requirement* of risk neutrality in well-being, to a rational *permission* to adopt other risk attitudes, does *not* salvage the compatibility of the EI veil of ignorance with *Minimal Preference for Equality*. Assume that risk neutrality in well-being is *one* of the permissible risk attitudes (as surely it should be). If there is a population of individuals, at least one of whose members has this attitude, then *that* individual will be indifferent between d^+ and d^{++} , and this will be rationally permitted. Given a plausible specification of the veil-of-ignorance framework for the case where individuals behind the veil have divergent evaluations of the distributions at issue,

²⁸ John Broome considers such a requirement, which he calls “Bernoulli’s hypothesis,” in *Weighing Goods* (Oxford: Basil Blackwell, 1995). See also Mattias Risse, “Harsanyi’s ‘Utilitarian Theorem’ and Utilitarianism,” *Nous* 36 (2002): 550-77.

²⁹ More formally, risk-aversion in well-being means maximizing the expectation of $f(w)$, where w is the well-being level of a bundle as assigned by the $w(\cdot)$ function, and $f(\cdot)$ is strictly concave. By Jensen’s Inequality, someone who sees d^{++} and d^+ as $1/N$ probabilities of their component bundles, and maximizes $f(w)$, will prefer d^{++} .

the upshot will be that d^+ and d^{++} are incomparably just.³⁰ But, surely, d^{++} is more just than d^+ —not incomparably just.

But why seek to undercut the veil-of-ignorance framework? Unfortunately, the veil of ignorance, at least on the EI construal, does not support the PD principle, for the same reason that it does not support *Minimal Preference for Equality*. A risk neutral deliberator behind the veil would be indifferent between a $1/N$ chance of being High in d and a $1/N$ chance of being Low in d (plus a $1-2/N$ chance of being anyone else in the population in d), as compared to a $1/N$ chance of being High in d^* and $1/N$ chance of being Low in d^* (plus a $1 - 2/N$ chance of being anyone else in the population in d^*).

Conversely, a famous theorem of Hardy, Littlewood, and Polya shows that any unequal distribution of some fixed total of a good can be transformed into a perfectly equal distribution via a series of Pigou-Dalton transfers.³¹ Thus the benefit-claim framework, if it supports the PD transfer (as I've argued it does), also supports Minimal Preference for Equality.

B. Temkin's Framework

In his influential book, *Inequality*, and subsequent writings, Larry Temkin has argued for the moral importance of “comparative fairness”³²: “[C]oncern about equality is a portion of our concern about fairness that focuses on how people fare relative to others Egalitarians in my sense generally believe that is bad for some to be worse off than others through no fault or choice of their own.”³³

Temkin discusses in detail how “comparative fairness” might be fleshed out as a criterion for ranking outcomes, as a function of individual well-being³⁴ and facts about individual responsibility. The ranking of outcomes is determined by individual “complaints.” Temkin's analysis of these “complaints” focuses on the simple case where individuals are not differentially

³⁰ On this specification, d^{++} is at least as just as d^+ iff all members of the population, behind the veil, weakly prefer d^{++} . This approach has the virtues of giving equal weight to the preferences of all members of the justice population, and of yielding a transitive (if incomplete) ranking of distributions.

³¹ See Albert Marshall and Ingram Olkin, *Inequalities: Theory of Majorization and its Applications* (New York: Academic Press, 1979), 21-22. The Hardy/Littlewood/Polya result is much more general, linking PD transfers and Lorenz dominance. To see why the proposition in the text is true: Let g be a list of the holdings of some good distributed unequally among the members of a finite population. Let $U(g)$ be the number of individuals whose holdings are *not* at the average level. There must be at least one below the average and one above. Pick any such pair, and let Δ be the lesser of the distance from the average to their two levels. Arrange a PD transfer of Δ between them, yielding g^* . Note that $U(g^*)$ is smaller (by a value of one or two) than $U(g)$. If this process is repeated, it must in a finite number of steps yield a vector with a U value of zero, i.e., perfect equality.

³² “Equality, Priority, or What?” *Economics and Philosophy* 19 (2003): 61-87, 62.

³³ Ibid. The book is *Inequality* (New York: Oxford University Press, 1996) See also Temkin's “Equality, Priority, and the Levelling Down Objection” and “Egalitarianism Defended,” *Ethics* 113 (2003): 764-82. Temkin does not use the term “comparative fairness” in *Inequality*, but the book is clearly focused on what he later comes to denote by that term. See *Inequality*, 13.

³⁴ “Throughout this work I shall mainly discuss inequality of *welfare*.” *Inequality*, 10.

responsible for their well-being levels.³⁵ In this case, whether someone has a complaint in a given outcome depends on her welfare level and the number and welfare levels of individuals who are better- and worse off than her.

Temkin considers different possibilities for identifying those who have complaints; for measuring the complaints' strength; and for comparing two outcomes in light of the pattern of complaints in each. One possibility is that each person in outcome x who is below the mean level of well-being in x has a complaint in x ; another, that each person in x other than the best-off person has a complaint in x ; a third, that each person in x other than the best-off person has *one or more* complaints in x , one for each of the persons better off than him. The strength of individual complaints might be measured in a linear or non-linear manner. Finally, one possible approach outlined by Temkin for comparing x and y in light of the patterns of complaints in each outcome is aggregative: x is fairer than y iff the total sum of complaints (or weighted complaints) in x is smaller than the total sum of complaints (or weighted complaints) in y . But Temkin also describes a non-aggregative, "maximin" rule for comparing two outcomes, namely: x is fairer than y iff the largest complaint in x is smaller than the largest complaint in y .³⁶

It seems no gross distortion of Temkin's account to see "comparative fairness" as a potential conception of *distributive justice*³⁷; and to adapt his model of "complaints" to the specific set-up described earlier. Thus adapted, Temkin's framework says: For a given distribution d^+ , an individual has zero, one, or more complaints. The number of her complaints and their strength potentially depends upon all the various features of d^+ , including her well-being level, the levels of other persons, and how prudently she and they behave. In turn, whether d^+ is more just than d^{++} depends upon the number and strength of complaints in d^+ , and the number and strength of complaints in d^{++} .

This "complaint" framework has both interesting similarities to, and critical differences from, the benefit-claim approach. A "claim," in the generic sense, is something *about* some particular person that is identified by a justificatory framework, and serves to organize thinking about distributive justice. This is a very natural device for deliberating about justice, and indeed *both* the benefit-claim and complaint frameworks employ that device. A Temkin-style "complaint" is one kind of "claim"; a Nagel-style benefit-claim is also one kind of "claim." But they differ, critically, in their structure. A Temkin-style "complaint" is a claim that takes the form of a relation between one person and one distribution; by contrast, a benefit-claim takes the form of a relation between one person and *two* distributions.

To say that some person (Sue, say) has a Temkin-style "complaint" is, more precisely, to identify one distribution d^+ and say: were d^+ to obtain, Sue would have a justified complaint, of a

³⁵ "[F]or each of this book's examples I shall assume that people are equally skilled, hardworking, morally worthy, and so forth, so that those who are worse off than others are so through no fault of their own." *Ibid.*, 17.

³⁶ See in particular *Inequality*, ch. 2, and also ch. 6 for a discussion of a non-linear measure of complaint strength.

³⁷ See, e.g., *Inequality*, 13.

certain strength, that d^+ is unfair to her. Thus, it is perfectly coherent to say, of two distributions, that Sue's complaints in one are stronger than her complaints in the second. By contrast, it would be incoherent to say that Sue has stronger *benefit-claims* in one distribution than a second. Rather, the structure of a benefit-claim is that it *compares* distributions from the perspective of one individual, giving her a claim of some strength in favor of one or other.

As a shorthand, we might say that a Temkin-style "complaint" has a "within distribution" structure, since it is a relation that embeds one distribution; while a benefit-claim has an "across-distribution" structure, since it embeds two distributions.

These divergent blueprints for the structure of claims, within-distribution versus across-distribution, are not only reasonably faithful adaptations (I believe) of Temkin's and Nagel's scholarship. They also correspond to divergent strands in ordinary discourse about fairness. Sometimes it is said to be fairest to reserve some decision, between two alternative courses of action, to those who are "affected" by it. Others need have no say, because they have no stake in the matter. This is to express an across-distribution view of claims: someone's claim, if she has it, will be a claim for one alternative *over* another, and she will have such a claim only if affected, i.e., better off with one *rather than* the other alternative.

However, sometimes those who are worse off than others are heard to complain that this inequality is unfair. The complaint is *about* the one outcome (here, the actual outcome) in which the inequality occurs. The worse-off are identifying a feature of the actual outcome (that there are others better off than them) which counts against it in a fairness comparison with *any* alternative, rather than identifying one particular counterfactual outcome relative to which they have grounds for complaint. This is to express a within-distribution view of claims.

This difference between the benefit-claim and "complaint" frameworks is not merely expositional: a matter of two alternative "stories" for the very same ranking of distributions in light of distributive justice. To start, it is far from clear whether the complaint framework supports the PD principle. Consider any case in which d and d^* meet the conditions stated by the PD principle, but Low is not the worst off person in these distributions. There are some individuals who are worse off than Low in d , and hence in d^* . Because High and perhaps others are better off than Low in d , she (Low) may well have complaints in d ; and the number and/or strength of Low's complaints may well diminish in d^* , since the difference between Low's welfare level and these individuals' diminishes in d^* . However, if Low has complaints in d , then surely also do the individuals in d who are worse off than Low (at least if they are equally or more prudent than Low and High). And these persons' complaints in d^* could increase in their number and/or strength; Low has become yet better off than them, and although High's welfare has decreased, he remains no closer to them than Low. Why believe that the most attractive specification of the "complaint" approach will *necessarily* identify d^* as more just?

Temkin himself has stated that the PD principle needs “serious modification,” writing: “[I]f the consequence of altering the gap between A and B is that the gaps between A and B and *other* groups are also altered, ... the net effect of such a transfer [on inequality] would depend, at least in part, on both the size and the number of *all* the different increases and decreases.”³⁸ In the margin, I show that some (if not all) of the methods outlined by Temkin for identifying complaints, measuring their strength, and comparing outcomes in light of their pattern can indeed lead to violations of the PD principle.³⁹

I would be delighted to be wrong about this—to be shown that the “complaint” framework supports the PD principle. (Defending the principle, not criticizing the framework, is the chief aim of this Article.) But if the “complaint” framework does not support the PD principle, then—in order to defend that principle—we need to undercut the framework. *Minimal Preference for Equality* here offers no help, since the “complaint” framework surely supports it.⁴⁰ Rather, I will appeal to a weak version of the Pareto principle, which runs as follows.

³⁸ *Inequality*, 83-84. There is also a subliterate on inequality metrics that fail to satisfy the PD principle—scholarship that is inspired, in part, by Temkin’s notion of complaints. See Chakravarty, *Inequality, Polarization and Poverty*, ch. 3.

³⁹ As mentioned, Temkin considers three possibilities with respect to identifying complaints, each with a corresponding reference level of well-being, in turn used to establish the strength (or equivalently, “size”) of complaints: (I) that everyone below the mean has one complaint (the reference level here is the mean); (II) that each person except the best off person has a complaint (the best-off person is now the reference level); and (III) that each person except the best off person has one or more complaints, namely one against each person better off than him (the reference level is each such person). In the initial chapters of *Inequality*, Temkin assumes that the strength of the complaint is just the difference between the reference level and the complainant’s well-being. Finally, he mentions four possibilities for comparing two outcomes in light of the pattern of complaints (for short, a “patterning rule”): additive, weighted additive (specifically, giving more weight to those with larger complaints), maximin, or a leximin variation on maximin (comparing the largest complaint in x to the largest in y , if those are equal the second-largest to the second-largest; etc.). It should be noted that a simple way to accomplish the weighted additive rule is to sum a convex transformation of the complaints.

It is clear that Possibility I in all variations violates PD: no one above the mean has complaints, and because the mean is the reference level PD transfers between individuals above the mean have no effect on the number or size of complaints. Possibility II violates PD with an additive patterning rule. For example, if we move from (10, 20, 100) to (15, 15, 100), the sum of complaints is 170 in both cases. Possibility II also violates PD with a maximin patterning rule: a PD transfer between individuals who are not worst off does not change the magnitude of the largest complaint. However, it can be shown that Possibility II with a leximin patterning rule, or summing a convex transformation of complaints, will satisfy the PD principle.

Possibility III is the most favorable for the PD principle. This approach even with an additive patterning rule satisfies PD, and also does summing a convex transformation, or leximinning complaints. It does not with a maximin rule.

Later, in chapter 6 of *Inequality*, Temkin argues that the difference rule for measuring the strength of complaints needs modification, but does not offer a concrete proposal for doing so. Regardless, Possibility I in all variants will continue to violate the PD principle with respect to transfers above the mean (unless such transfers, even though mean-preserving, somehow change the strength of complaints held by those below the mean). And all possibilities coupled with the maximin patterning rule will violate the PD principle.

⁴⁰ This is true for reasons discussed several paragraphs below in the analysis of the Minimal Pareto Principle.

Minimal Pareto Principle. Let d^+ and d^{++} be such that everyone is better off in d^{++} than d^+ , and everyone is sufficiently prudent. Then: d^{++} is more just than d^+ .

The term “Pareto” is sometimes used to name a principle concerning preferences, sometimes a principle about well-being. The version of the Pareto principle just expressed is the latter sort. It is weak in three ways. First, unlike the so-called “strong” Pareto principle (in terms of well-being) which comes into play when *some* are better off with one alternative and everyone else is at least as well off, the Minimal Pareto Principle constrains the ranking of distributions only if *everyone* is better off.

Second, it expresses only a *pro tanto* principle. The Pareto principle is typically formulated as an all-things-considered principle, namely that alternative⁴¹ a^{++} is all-things-considered morally better than alternative a^+ if some are better with a^{++} than a^+ and everyone is at least as well off, or alternatively (more weakly) if everyone is better off with a^{++} than a^+ . But these all-things-considered principles can be accused of anthropocentrism: imagine that all persons are better off with a^{++} , but a^+ is massively worse for entities with moral significance that are not persons. The all-things-considered principles can also be challenged for ignoring considerations of *desert*—a point stressed by Temkin,⁴² and implicit in Shelly Kagan’s recent pathbreaking scholarship on desert.⁴³ Imagine that everyone is better off with a^{++} , but some receive more than they noncomparatively deserve, while in a^+ everyone gets exactly what she noncomparatively deserves. Then a^{++} will be worse than a^+ in light of noncomparative and perhaps also comparative desert, and could be worse than a^+ all things considered if desert has sufficient moral weight.⁴⁴

This is not to say that the all-thing-considered Pareto principles are *wrong*, just that they are controversial by virtue of their ambitious implication that the totality of Pareto-respecting moral considerations outweigh the totality of those that are not Pareto-respecting. The Minimal Pareto Principle makes no such claim, since it concerns *pro tanto* (not all-things-considered) moral betterness in light of one moral factor, distributive justice. And now note the third way in which the principle is weak: it builds in a responsibility requirement, stipulating that distributive justice favors the distribution that benefits everyone *if* everyone has been sufficiently prudent.

⁴¹ “Alternative” is meant to be generic, including outcomes, choices, or (as in this Article) distributions.

⁴² See “Equality, Priority, and the Levelling Down Objection” 138-40.

⁴³ Kagan, *The Geometry of Desert*.

⁴⁴ The analysis here is not meant to *endorse* desert as a moral consideration separate from justice (Rawls famously denied that it was), but to be agnostic on the issue. If desert is a separate such consideration, its relation to justice understood along responsibility-adjusted-welfarist lines is complex. The features of an individual’s character and choices that ground her degree of responsibility, as opposed to her degree of desert, could well diverge. Thus the Minimal Pareto Principle (even modified to allow non-prudence bases for responsibility) might well favor a distribution worse in light of desert.

The benefit-claim framework readily supports the Minimal Pareto Principle. Conceivably individuals who have been badly imprudent lack any claim in distributive justice that one distribution rather than a second obtain, even if they are affected by which does. But we are considering a case in which everyone is not only better off in d^{++} than d^+ , but also sufficiently prudent. Thus everyone has a benefit-claim valenced in favor of d^{++} . All the considerations on the table (parceled out among the justice population in the form of benefit-claims) point in favor of d^{++} ; none point in favor of d^+ . In such a case, the benefit-claim framework surely prefers the distribution that is universally favored. The point of this justificatory framework (like any other) is to bring to light how various facts about the world and persons interrelate to constitute some aspect of moral rightness. If d^{++} and d^+ are as specified by the Minimal Pareto Principle, *all* the factors identified by the framework as morally relevant—all the benefit-claims—point in one direction, towards d^{++} . How could d^{++} be anything but more just than d^+ ?

By contrast, under the “complaint” framework, d^{++} will be less just than d^+ in *many* cases, and indeed in *every* case of the following kind: Everyone is better off in d^{++} than d^+ and sufficiently prudent (so the Minimal Pareto Principle favors d^{++}) but (1) everyone is at the same well-being level in d^+ , while well-being is unequally distributed in d^{++} ; and (2) everyone is equally prudent. *Whatever* the rule for identifying those with “complaints” and measuring the complaints’ strength, it must surely be the case that none have complaints in d^+ , and some have complaints in d^{++} . No one is at comparative fault, and no one is worse off than anyone else in d^+ , while some *are* worse off than others in d^{++} . And, *whatever* the rule for ranking two distributions in light of the pattern of complaints in each, surely d^{++} is less just (according to the “complaint” framework) than d^+ . All the moral factors identified by *that* framework—all the “complaints”—count against d^{++} . If d^{++} were to obtain, some would have cause for complaint in the Temkin sense; if d^+ were to obtain, none would.

Thus the “complaint” framework would rank distributions in conflict with the Minimal Pareto Principle; and for that reason, I suggest, we should reject the framework. The Minimal Pareto Principle is extremely plausible. The principle, it should be stressed, is not vulnerable to Temkin’s challenge to the stronger and more familiar versions of the Pareto principle: namely, that alternatives which make everyone worse off might still be morally better, because they implicate *impersonal* moral values.⁴⁵ While there may well be impersonal moral value in the flourishing of non-human animals, in giving persons what they morally deserve, in beauty rather than ugliness, or in creating persons who otherwise would not exist, distributive justice is a quintessentially *personal* moral consideration. A distribution of goods is *fair* only if it respects the separateness of persons: only if it takes separate account of each persons’ interests, concerns, goals, or values, and can be seen as justifiable on balance to each individual, given her interests, concerns, goals, or values. This insight, of course, goes back to Rawls; and the veil-of-ignorance framework, as well as Temkin’s and Nagel’s, are all attempts to structure

⁴⁵ See “Equality, Priority, or What?” 77; “Equality, Priority and the Levelling Down Objection.”

distributive assessment so as to secure such respect. They are all (as it were) specific *interpretations* of the “separateness of persons.”

But if one alternative is better than a second in light of each person’s interests, then it is justifiable to each, and the second justifiable to none. Thus a concern for *fair distribution* favors the first alternative, whatever other moral factors and morality on balance might say. It might be objected that “respecting the separateness of persons” means attending to each person’s goals, values, concerns, or something else about her *other than* her interests (well-being). But the welfarist about distributive justice (responsibility adjusted or not) will hardly find *this* objection to be persuasive. Welfarism about distributive justice just *is* the view that the right way to respect each person is to attend to her *interests*, and thus that the pattern of individuals’ well-being is what determines how well justice has been achieved.

The argument here for the Minimal Pareto Principle is, therefore, an argument *within* welfarism about distributive justice. In that sense, the argument is modest; the principle (even though minimal) may well be harder to argue for if some non-welfarist currency is adopted. Still, the argument reaches a non-trivial, even surprising conclusion.⁴⁶ The within-distribution view of claims turns out to be a *failed* interpretation of the separateness of persons, given the welfarist view about which aspect of a person (on that view, her well-being) demands consideration.

Finally, a word might be said about how this argument relates to the literature on “leveling down.”⁴⁷ Plausible principles of distributive justice will involve a preference for equalization of the goods (whether welfare or something else) covered by those principles. But it is possible (whether or not plausible) that equality is *also* one of the impersonal moral values. It may be better “from the point of view of the universe” for a more rather than less equal pattern of well-being, or something else, to obtain. If so, “leveling down”—making everyone worse off so as to reduce inequality—will be morally better *pro tanto*, in light of that value, and perhaps all-things-considered. But *distributive justice* does not counsel leveling down, or so I have argued. If there is moral value in pursuing equality by leveling down, then that value is separate from distributive justice.

IV. *Prioritarianism*

Prioritarianism, often, is understood a kind of welfare-consequentialism, namely an account of the moral ranking of outcomes as determined by the well-being of individuals therein. But what are the axiomatic features of this ranking? It seems clear that prioritarians understand it to satisfy the Pigou-Dalton principle for outcomes, namely:

⁴⁶ Temkin, surely, would be surprised, since he articulates the “complaint” framework in *Inequality* using differences in *well-being* as the basis for individuals’ complaints.

⁴⁷ See scholarship cited in Adler, *Well-Being and Fair Distribution*, 337, n. 53.

PD Principle for Outcomes: If High is better off than Low in outcome x , and at least as well off as Low in outcome y ; High is better off in x than y , Low in y than x , with the difference in High's well-being between x and y exactly equal to the difference in Low's well-being between y and x ; and everyone else is equally well off in the two outcomes, then: y is a morally better outcome than x .

For example, in the canonical statement of prioritarianism, Parfit writes: “[b]enefitting people matters more the worse off these people are”; that well-being has “diminishing marginal moral importance”; and that: “For Utilitarians, the moral importance of each benefit depends only on how great the benefit would be. For *Prioritarians*, it also depends on how well off the person is to whom this benefit comes. We should not give equal weight to equal benefits, whoever receives them. Benefits to the worse off should be given more weight.”⁴⁸ All these formulations (in the context of morally ranking outcomes) point to the PD principle for outcomes.

However, Parfit also stresses that prioritarians are not concerned with “relativities.” The moral weight of a change to someone’s well-being does not depend upon how her level of well-being compares with others.

People at higher altitudes find it harder to breathe. Is this because they are higher up than other people? In one sense, yes. But they would find it just as hard to breathe even if there were no other people lower down. In the same way, on the Priority View, benefits to the worse off matter more, but that is only because these people are at a lower *absolute* level. It is irrelevant that these people are worse off *than others*. Benefits to them would matter just as much even if there *were* not others who were better off.

The chief difference, then, is this. Egalitarians are concerned with *relativities*: with how each person’s level compares with the level of other people. On the Priority View, we are concerned only with people’s absolute levels.⁴⁹

If prioritarianism does indeed have the feature that the moral weight of a change to someone’s well-being is independent of how her well-being compares with others, then the prioritarian ranking of outcomes should surely satisfy an axiom of “separability.”

Separability for Outcomes: The moral ranking of two outcomes is independent of the well-being levels of unaffected individuals (those who are equally well off in the two outcomes).

Assume that there are A individuals not equally well off in x and y ; and U unaffected individuals. Each of the affected individuals has some well-being level in x , and a different level in y . The prioritarian ranks x and y by assigning each such welfare difference a moral value—a value that is *just* a function of the individual’s levels in the two outcomes, and not of how her well-being compares to others. And the comparison of x and y then depends upon these A moral values.

⁴⁸ Parfit, “Equality or Priority,” 101, 105.

⁴⁹ *Ibid.*, 104.

What those A values are, and how x and y compare in light of them, has nothing to do with what the well-being levels of the U unaffected individuals happen to be.

In short, separability seems to be one of the axioms for ranking outcomes that prioritarrians should endorse, in virtue of what they have said about the non-relativistic nature of prioritarianism. Indeed, in the literature on social choice, prioritarianism is almost always expressed as a social welfare function that satisfies *both* the PD principle for outcomes and separability.⁵⁰

What is the logical connection between the PD principle for outcomes and separability? None. A moral ranking of outcomes can satisfy the PD principle but not separability; satisfy separability but not the PD principle; both; or neither. The last three cases are illustrated in the margin.⁵¹ The most important case, for our purposes, is the first. Consider the “rank-weighted” rule for ordering outcomes. With N individuals in the population, N different fixed positive weights are specified, $w_1 > w_2 \dots > w_N > 0$. A given outcome is assigned a number, equaling the well-being level of the lowest-ranked (i.e., worst-off) individual multiplied by w_1 , plus the well-being level of the second-worst-off individual multiplied by w_2 , ... plus the well-being level of the best-off individual multiplied by w_N .⁵² Outcomes are then ordered according to these sums.

The rank-weighted approach *satisfies* the PD principle for outcomes. This is obvious in the scenario where the transfer does not cause anyone to change ranks. For example, imagine that there are three individuals, Jim, Ken, and Larry. In outcome x , their well-being levels are, respectively, 20, 50, and 200, assigned a numerical value of $20w_1 + 50w_2 + 200w_3$. A PD transfer occurs, transferring 15 units from Larry to Ken. The numerical value assigned to this new outcome is $20w_1 + (50+15)w_2 + (200-15)w_3$. Since $w_2 > w_3$, this numerical value must be greater than the numerical value assigned to x .

⁵⁰ John Broome, “Equality or Priority: A Useful Distinction” (n.d.), available at <http://users.ox.ac.uk/~sfop0060/pdf/equality%20versus%20priority.pdf>; Campbell Brown, “Priority or Sufficiency ... or Both?” *Economics and Philosophy* 21 (2005): 199-220; Marc Fleurbaey, “Equality versus Priority: How Relevant is the Distinction?” (2001); Iwao Hirose, “Reconsidering the Value of Equality,” *Australasian Journal of Philosophy* 87 (2009): 310-12; Nils Holtug, “Prioritarianism,” in *Egalitarianism*, ed. Nils Holtug and Kasper Lippert-Rasmussen (Oxford: Clarendon Press, 2007), 125-56; Karsten Klint Jensen, “What is the Difference Between (Moderate) Egalitarianism and Prioritarianism?” *Economics and Philosophy* 19 (2003): 89-109; David McCarthy, “Utilitarianism and Prioritarianism II” *Economics and Philosophy* 24 (2008): 1-33; Wlodek Rabinowicz, “Prioritarianism and Uncertainty,” in *Exploring Practical Philosophy*, ed. Dan Egonsson et al. (Aldershot: Ashgate, 2001), 139-65; Bertil Tungodden, “The Value of Equality,” *Economics and Philosophy* 19 (2003): 1-44.

⁵¹ The utilitarian ranking of outcomes, summing individuals’ well-being levels, is separable but does not satisfy the PD principle. The prioritarian ranking of outcomes—summing a strictly concave transformation of individuals’ well-being levels—satisfies both. A rank-weighted rule with weights that are not strictly decreasing will satisfy neither the PD principle nor separability.

⁵² This statement of the rank-weighted rule, and the examples in the footnote immediately above, assume that each person’s well-being in some outcome is represented by a single number; but the analysis readily generalizes to the case in which a set of numerical well-being values are assigned to a given pairing of an individual and outcome (a numerical representation scheme that allows for well-being incomparability). See Adler, *Well-Being and Fair Distribution*, ch. 5.

For a demonstration that the rank-weighted approach satisfies the PD principle for outcomes in general (even when transfers cause rank-switches), see the Appendix.

However, the rank-weighted approach *violates* separability. The moral weight of someone's well-being difference as between two outcomes depends upon his well-being rank in the entire population—not just the two levels taken alone. To illustrate, let outcome y be such that Jim still has 20—so he is unaffected—but Ken has 60, and Larry 175. Thus while x is assigned a numerical value of $20w_1 + 50w_2 + 200w_3$, y is assigned a value of $20w_1 + 60w_2 + 175w_3$. The difference between these values is $w_2(60-50) + w_3(175-200)$.

Now consider two outcomes x^+ and y^+ which are identical to x and y , respectively, except that the unaffected person (Jim) is at a different level of well-being. He remains unaffected between x^+ and y^+ , but his position in the population pattern of well-being is different. For example, x^+ is such that Jim has 100, Ken 50, and Larry 200, while y^+ is such that Jim has 100, Ken 60, and Larry 175. The rank-weighted approach assigns x^+ a value of $w_2(100) + w_1(50) + w_3(200)$, and y^+ a value of $w_2(100) + w_1(60) + w_3(175)$. The difference between *these* values is $w_1(60-50) + w_3(175-200)$. Note that *this* numerical difference could be positive while the numerical difference between y and x negative—since $w_1 > w_2$. A violation of separability can occur because in the x^+/y^+ comparison, Ken's well-being difference is multiplied by w_1 (he is the lowest ranked in both of these outcomes), while in the x/y comparison it is multiplied by w_2 (he is the second-lowest-ranked).

The rank-weighted approach to ranking outcomes is a PD-respecting approach, but violates separability, and thus is not a *prioritarian* approach. Now, self-described “prioritarians” might object that I have mischaracterized their commitments. This is an interpretive question which need not be belabored here. The key observation is that nothing in the logic of rankings *requires* the combination of separability and the PD principle. A substantive argument is needed to show why we should endorse this combination—rather than the PD principle without separability—as part of the criterion for ranking outcomes.

Let us turn now to justice. Recall that we are focusing throughout on principles for ordering subsets of distributions with fixed prudence. Within any such subset, what varies about individuals (if anything does) is their well-being; and separability for distributions (for short, “separability”) is straightforwardly definable by direct transposition from the outcome case.

Separability for Distributions: The justice ranking of two distributions does not depend upon the well-being levels of unaffected individuals.

Once more, we need a substantive argument for coupling *this* principle with the PD principle (for distributions). Does the benefit-claim framework furnish such an argument?

I believe it does, but I also believe that this argument is not nearly as straightforward as the argument for the PD principle itself. Recall the structure of the argument for the PD

principle: the facts about High's and Low's levels furnish a pro tanto reason for the distribution (d^*) better for Low; and that consideration is not overridden by any other. Notably, this argument conceded the possibility that the strength of an affected person's benefit-claim *might* depend not only upon her well-being and responsibility levels in the two distributions, but also upon the well-being levels of everyone else in the population. Even if well-being rank *were* a factor influencing the strength of someone's benefit-claim, that factor (so I argued) simply reinforced Low's claim at the expense of High's.

By contrast, the argument from the benefit-claim framework to separability must show that the methodology for assigning strength to a non-zero claim cannot make reference to the claimant's well-being rank in the distributions being compared, or otherwise take account of the well-being levels of unaffected individuals. But why not? Do not be misled by the "across-distribution" architecture of benefit claims. Such a claim (by contrast with a Temkin "complaint") is a relation between an individual and two distributions; but its strength can depend upon *any* facts about the two distributions, including how her well-being compares to others' in each of the two.

For example, consider the following rule (fragment)⁵³ for assigning strength to benefit claims. If the two distributions being compared are such that each individual's rank in the population distribution of well-being does not vary between the two, let the strength of each affected person's claim (in favor of the distribution where she is better off) be 1 multiplied by her well-being difference if she has the highest level of well-being, 2 multiplied by her well-being difference if she has the second-highest level, etc; and then sum these weighted claims to compare the distributions. This can yield a violation of separability, as the following example shows: d is preferred to d^* but d^{++} to d^+ , because shifting the level of well-being of unaffected Jim shifts Ken's rank from second-highest to third-highest, and thus changes the strength of Ken's benefit-claim.

| | <u>Jim</u> | <u>Ken</u> | <u>Larry</u> |
|---------------------------|------------|------------|--------------|
| Distribution d | 20 | 50 | 200 |
| Distribution d^* | <u>20</u> | <u>60</u> | <u>175</u> |
| Strength of benefit-claim | 0 | 2(10) | 1(25) |
| Distribution d^+ | 100 | 50 | 200 |
| Distribution d^{++} | <u>100</u> | <u>60</u> | <u>175</u> |
| Strength of benefit-claim | 0 | 3(10) | 1(25) |

⁵³A full rule would also cover cases in which some individuals change rank. In order to illustrate how the benefit-claim framework *might* violate separability, it suffices to show that a rule-fragment of the sort here described would do so.

In short, it is the PD principle (together with the Minimal Pareto principle) that responsibility sensitive welfarists should see as the core of justice. While a plausible and, I believe, persuasive case *can* be made for “prioritarianism” about justice—for adding separability—that case is certainly less compelling than the case for the PD and Minimal Pareto principles. These are the twin pillars of justice.

V. Appendix

This Appendix explains why the rank-weighted approach to ordering outcomes satisfies the PD principle for outcomes (see above Part IV).

Let $\mathbf{u}(x) = (u_1(x), u_2(x) \dots, u_N(x))$ be a list of each person’s well-being in outcome x , with u_1 the well-being of person 1 in x and so forth. Let $\tilde{\mathbf{u}}$ be a permutation (rearrangement) of the entries in $\mathbf{u}(x)$, with $\tilde{u}^1(x)$ the first entry, and which is “rank ordered,” i.e., $\tilde{u}^1(x) \leq \tilde{u}^2(x) \leq \dots \leq \tilde{u}^N(x)$. The rank-weighted approach to ordering outcomes uses fixed positive weights $w_1 > w_2 > \dots > w_N > 0$, and assigns x a value $r(x)$ equaling: $w_1\tilde{u}^1(x) + w_2\tilde{u}^2(x) + \dots + w_N\tilde{u}^N(x)$. Outcomes are then ordered according to the rule: x is morally at least as good as y iff $r(x) \geq r(y)$.

Assume that y is reached from x via a PD transfer from individual h (igh) to individual l (ow). That is, $\Delta > 0$; $u_l(x) < u_l(y) = u_l(x) + \Delta \leq u_h(y) = u_h(x) - \Delta$; and for all $k \neq h, l$, $u_k(x) = u_k(y)$. Let us say that y and x are “rank consistent” (in the case of a PD transfer) if there is no k such that $u_k(x) > u_l(x)$ but $u_k(y) < u_l(y)$ and no k such that $u_k(x) < u_h(x)$ but $u_k(y) > u_h(y)$.

It is easy to see that the rank-weighted rule gives a higher value to y over x if the outcomes are rank consistent. Note that this is clearly true even if individual l , or h , or both starts out or ends up at a level equal to some other person. For example, if the well-being levels in x are (1, 2, 6, 13, 17), and the individual at level 17 moves to 13, while the individual at level 2 moves to 6—resulting in well-being levels in y of (1, 6, 6, 13, 13)—the value $r(x)$ is $w_1(1) + w_2(2) + w_3(6) + w_4(13) + w_5(17)$. The value $r(y)$ is $w_1(1) + w_2(6) + w_3(6) + w_4(13) + w_5(13)$. Since $w_5 < w_2$, $r(y) > r(x)$.

If y and x are not rank-consistent, we can create a series of rank-consistent PD transfers from x to m to m^* to ... to y , such that each outcome in this series is assigned a lower $r(\cdot)$ value than the next, and thus (by the transitivity of the real numbers) $r(x)$ must be less than $r(y)$. For example, imagine that x as before has the well-being values (1, 2, 6, 13, 17), and there is a transfer of 6 units from the last individual to the first—so that y has the values (7, 2, 6, 13, 11). Let m be (2, 2, 6, 13, 16); m^* (5, 2, 6, 13, 13); and m^{**} (6, 2, 6, 13, 12). Then x to m , m to m^* , m^* to m^{**} , and m^{**} to y are each rank-consistent PD transfers.

To construct a general rule for doing this, assume that there are K individuals who are strictly better off than l in x but worse off than her in y ; and M individuals who are strictly worse off than h in x but better off than her in y . Denote the well-being levels of the K individuals as R_1, \dots, R_K and the well-being levels of the M individuals as S_1, \dots, S_M . Let u_l be the well-being

level of individual l in x , u_h the level of individual h in x , and $u_l + \Delta$ and $u_h - \Delta$, respectively, their levels in y .

Let Δ_1 be the minimum positive value in the set of numbers $\{R_1 - u_l, R_2 - u_l, \dots, R_K - u_l, u_h - S_1, u_h - S_2, \dots, u_h - S_M\}$. Let Δ_2 be the minimum positive value in the set of numbers $\{R_1 - (u_l + \Delta_1), R_2 - (u_l + \Delta_1), \dots, R_K - (u_l + \Delta_1), (u_h - \Delta_1) - S_1, (u_h - \Delta_1) - S_2, \dots, (u_h - \Delta_1) - S_M\}$. Let Δ_3 be the minimum positive value in the set of numbers $\{R_1 - (u_l + \Delta_1 + \Delta_2), R_2 - (u_l + \Delta_1 + \Delta_2), \dots, R_K - (u_l + \Delta_1 + \Delta_2), (u_h - \Delta_1 - \Delta_2) - S_1, (u_h - \Delta_1 - \Delta_2) - S_2, \dots, (u_h - \Delta_1 - \Delta_2) - S_M\}$. Repeat this process, recursively, until a Δ_T is reached such that no positive Δ_{T+1} exists.

Now consider the series of T outcomes m_1, \dots, m_T such that individual l in outcome m_t has well-being level $u_l + \Delta_1 + \Delta_2 + \dots + \Delta_t$, while individual h has well-being level $u_h - \Delta_1 - \Delta_2 - \dots - \Delta_t$ (while all other individuals have their well-being levels in x , equaling their levels in y). By construction, each outcome in the series x, m_1, \dots, m_T, y is reached via a rank-consistent PD transfer from the one before.

Note: it can also be proved by algebraic manipulation that the rank-weighted approach satisfies the PD principle; but this method of decomposing any PD transfer into a series of rank-consistent PD transfers shows, quite intuitively, why that is the case.