**Weighing and Aggregating Reasons under Uncertainty: a Trilemma[[1]](#footnote-1)\***

*Abstract*

*I discuss the trilemma that consists of the following three principles being inconsistent:*

*1. The Common Principle: if one distribution, A, necessarily brings a higher total sum of personal value that is distributed in a more egalitarian way than another distribution, B, A is more valuable than B.*

*2. (Weak) ex-ante Pareto: if one uncertain distribution, A, is more valuable than another uncertain distribution, B, for each patient, A is more valuable than B.*

*3. Pluralism about attitudes to risk (Pluralism): the personal value of a prospect is a weighted sum of the values of the prospect’s outcomes, but the weight each outcome gets might be different from the probability the prospect assigns to the outcome.*

1. *Introduction*

Distributive justice has two related but distinct aspects, a structural aspect and a substantive one. The structural aspect has to do with the question of how to distribute whatever it is which is the appropriate object of distribution. For convenience I will use the term “personal value” to refer to this object. The substantive aspect has to do with the question of what the appropriate object of distribution is.

In this paper I present a trilemma between two structural principles of distributive justice and one substantive principle. Although the last is a substantive principle, it does not point to a specific interpretation of personal value as the correct one. Rather, it points to a feature that any interpretation of personal value must, prima facie, have (but, of course, since all the three principles of the trilemma seem, prima facie, justified, first impressions do not account for much in this paper). Specifically, it is a permissive principle regarding the structure of personal value.

I will not argue for any specific solution to the trilemma (a “solution” being a rejection of one of its the three horns). Adopting different interpretations of personal value (e.g. welfare, utility, capabilities, basic goods, economics resources) should plausibly lead one to accept different solutions. However, I believe that the trilemma is interesting when considered with respect to each one of the central interpretations of personal value found in the literature and the conclusions I will argue for are general enough to apply to all (or at least most) of those interpretations.

There are two general conclusions I will argue for. The first concerns the relation between consequentialist and non-consequentialist considerations in distributive justice. I will argue that one way to interpret the trilemma is as a demonstration of how these two types of consideration might be in very strong tension in some cases (stronger and different in nature, I believe, than the tension that is sometimes recognized in the literature). This general conclusion can be translated to specific objections to specific positions in the distributive justice literature that combine (sometimes implicitly) consequentialist and non-consequentialist elements. I will demonstrate this with respect to a recent argument by Lara Buchak (2017).

The second conclusion is that there is an intimate and surprising connection between seemingly independent questions – the question of how to weigh reasons for action under uncertainty and the question of whether or not the interest of groups, as groups (that is, independently of the interests of the groups’ members), matter in the context of distributive justice.

The paper is organized in the following way. In section 1 I present the trilemma, demonstrate it using a simple example and locate it within the existing literature. In sections 2-4 I discuss each horn of the trilemma in turn and point to potential philosophical implications of rejecting each one of these horns. In section 5 I conclude.

1. *The trilemma*

The trilemma consists of the mathematical fact that there are cases in which the following three principles are in conflict:[[2]](#footnote-2)

*1. The Common Principle*: if one distribution, A, *necessarily* (i.e. in every state) brings a higher total sum of personal value that is distributed in a more egalitarian way than another distribution, B, A is more valuable than B.[[3]](#footnote-3)

*2. (Weak) ex-ante Pareto:* if one uncertain distribution, A, is more valuable than another uncertain distribution, B, for each patient, A is more valuable than B.

*3. Pluralism about attitudes to risk (Pluralism)*: the personal value of a prospect is a weighted sum of the values of the prospect’s outcomes, *but the weight each outcome gets might be different from the probability the prospect assigns to the outcome.*[[4]](#footnote-4)

All three principles are discussed in depth in the next sections. However, a few brief remarks regarding their status in the contemporary literature are warranted here.

Although I am unaware of any explicit formulation of *The Common Principle* in the literature, it is easy to see that it is weaker than, and implied by each of the main positions in the literature regarding the structure of distributional justice: Utilitarianism (understood as a thesis about the structure of distributive justice, not as a thesis about the object of distributive justice), Egalitarianism and Prioritarianism. It is also highly intuitive, of course. In population ethics, it is often made use of in different versions of Parfit’s (1984) argument for the Repugnant Conclusion (usually without being explicitly formulated as a principle).[[5]](#footnote-5) The only other direct mention of an instance I am aware of is in Broome (1987), which is discussed below.

Although few philosophers reject *ex-post Pareto* (i.e. Pareto over certain distributions),[[6]](#footnote-6) the *ex-ante Pareto* principle is more controversial. Fleurbaey and Voorhoeve (2013), for example, explicitly argue that in some cases it should be violated (their argument is discussed in section 3). However, even those who reject it seem to assume that violations can be justified only in special circumstances and that, if possible, it should be obeyed (see for example Fleurbaey 2010). Among economists, the ex-ante principle is widely accepted (although not universally so, as the Fleurbaey example shows; see also Mongin 2016).

*Pluralism* is usually discussed in the context of decision theory. In that context, i.e. when personal value is understood to be utility in the sense of a representation of rational preferences, *Pluralism* is inconsistent with the demands of orthodox decision theory. Orthodox decision theory (as formulated by von Neuman and Morgenstern 1947 and Savage 1954) demands that the value of a prospect is equal to the prospect’s expected value; that is, to the weighted average of the values of the outcomes, when the weight each outcome gets is equal to its probability according to the prospect. Henceforth, I will refer to any value function that respects this demand as one that has an “expectational form”.

The demand that utility will have an expectational form is known to be problematic from a descriptive point of view.[[7]](#footnote-7) However, recently several philosophers (notably, Buchak 2013 and Stefansson and Bradley 2015 and 2017; see also Goldschmidt and Nissan-Rozen, 2020) have argued that it is too restrictive also from a normative point of view. Different scholars relax the demand in different ways. However, the trilemma seems to arise in the context of at least most of the alternatives to orthodox decision theory discussed in the literature. To avoid unnecessary technicalities, the discussion will proceed by concentrating on Buchak’s theory, which is the theory most discussed by philosophers in recent years. However, below I show that problem arises for a very wide family of deviations from orthodox decision theory.[[8]](#footnote-8)

I am not aware of any explicit discussion of *Pluralism* in the context of other interpretations of personal value. As will be discussed in the next section, however, it seems to me that the demand that personal value will have an expectational form is even less plausible under these interpretations.

A simple example, summarized in table 1, demonstrates the trilemma. There are two patients, two possible states and two possible uncertain distributions. Both patients assign probability 0.5 to each state. This probability distribution can be understood as subjective or objective. Nothing in the discussion depends on this. The first number in each cell designates the personal value in this cell for patient 1 and the second number designates the personal value in the cell for patient 2.

|  |  |  |
| --- | --- | --- |
|  | State 1 (probability 0.5) | State 2 (probability 0.5) |
| Distribution A | 0, 2.7 | 7, 2.7 |
| Distribution B | 1.9, 1.9 | 4.9,4.9 |

*Table 1*

Assume patient 1 assigns equal weights to both states (in line with the prescriptions of orthodox decision-theory). Thus, her personal value from A is 0.5\*0 + 0.5\*7 = 3.5, which is higher than her personal value from B, which is 1.9\*0.5 + 4.9\*0.5 =3.4.

Assume also, however, that patient 2 – in line with *Pluralism* – assigns a higher weight to state 1 (maybe because he disvalues risk). In such a case, even though B’s expected personal value for patient 2 (which is 1.9\*0.5 + 4.9\*0.5 = 3.4), is higher than A’s expected personal value (which is 2.7), A might be more valuable for him than B. This might be the case, for example, if patient 2 obeys the decision theory suggested by Buchak (2013) and has moderately risk avoidant preferences.

Buchak’s theory demands that the value of a prospect is a weighted average of the values of the prospect’s outcomes when the weights are given by a function, r(.), which assigns to the probability of each possible outcome a number between 0 to 1, according to how high the given outcome is ranked in the patient’s preferences. The more convex this function the more risk avoidant the patient is.

Let us assume that the r(.) function of patient 2 is r(p) = p2. In such a case, according to Buchak’s theory, while the value of A for patient 2 is 2.7, the value for him of B is 1.9 + 0.25\*3 = 2.65. It is easy to explain the rationale behind this calculation even without specifying the exact functional form that Buchak suggests: in B, patient 2 gets 1.9 units of personal value for certain and with probability 0.5 he gets an extra 3 units of personal value. Due to his r(.) function, the weight patient 2 gives to the latter possibility is not equal to its probability – which is 0.5 – but rather to the square of its probability, i.e. to 0.25.[[9]](#footnote-9) However, Buchak’s theory is not the only theory that allows for cases in which A is more valuable for patient 2 than B. The same goes for many other theories that accept *Pluralism*.

Thus, for both patients, A is more valuable than B. However, it is easy to see that in both states, B brings both a higher total sum of personal value and a more egalitarian distribution of this total sum. Thus, given *Pluralism*, either *ex-ante Pareto* or *The Common Principle* must go.

As an anonymous referee commented, a natural reaction to this example is to try and restrict the set of permissible weights the patients can use in a way that will prevent the conflict from arising. However, there exists no such plausible restriction (except maybe the one discussed below in footnote 10). It is possible to build an example with the same structure as the example above and in which patient 1’s personal value has an expectational form, in which the conflict arises, *for any assignment* of positive weights to the states by patient 2. Since all decision theories (that I am aware of) *allow* the patients to use value functions with an expectational form, this shows that no such plausible restriction on the weights (that still allows for different patients to use different weights) exits.

To see that consider the generalization of the example presented in table 2

|  |  |  |
| --- | --- | --- |
|  | State 1 (probability 0.5) | State 2 (probability 0.5) |
| Distribution C | 0, X | Z, Z-X-3Y |
| Distribution D | $\frac{X+Y}{2}$, $\frac{X+Y}{2}$ | Z -$ \frac{X}{2}$ - Y, Z -$ \frac{X}{2}$ – Y |

*Table 2*

Assume

1. Z > X+3Y > X > Y > O.
2. The personal values of the prospects for Patient 1 have an expectational form (i.e. patient 1 assigns a weight of 0.5 to both states).
3. Patient 2 assigns to state 1 a (positive) weight, w, and to state 2 a (positive) weight, p.

It is easy to verify that:

1. In both states D brings both a higher total sum of value and a more egalitarian distribution of this total sum.
2. C is more valuable than D for patient 1.

What about patient 2?

For C to be more valuable than D for patient 2, it must be that

wX + p(Z-X-3Y) > w($\frac{X+Y}{2}$) + p$(Z-\frac{X}{2}$- Y)

and from the above equation it is straightforward to get:

$\frac{w}{p}$ > $\frac{X+4Y}{X-Y}$

As Y goes to 0 the righthand expression goes to 1 so for any two positive numbers, w, p, such that w>p, we can find a positive Y small enough to make it the case that C is more valuable than D for patient 2 too.

This shows that even in cases with the simple structure we have here, i.e. when there are only two states with equal probability and two patients that the personal values function of one of them has an expectational form, it is possible to build an example in which *The Common Principle* clashes with *ex-ante Pareto* for any assignment of weights to the two states by the other patient, as long as these weights are not equal.

From a formal point of view, the trilemma is closely related to a well-known mathematical phenomenon, the ethical implications of which are discussed by Broome (1987)[[10]](#footnote-10). Formulated in the terminology introduced above, the phenomenon is as follows: When the personal value functions of all patients have an expectational form, but different patients assign different probabilities to the different states (that is they disagree about the probabilities a prospect assigns to its different outcomes), *ex-ante Pareto* is inconsistent with *The* *Common Principle.*

Broome uses this phenomenon to argue that the appropriate object of distributive justice (what he calls “personal good”) must be such that it takes as input the same probability function for all patients. When the same probability function is used in calculating the personal value of all patients, and when the personal values of all patients have an expectational form, the two principles are consistent. This, for example, (according to Broome) rules out utility, understood as a representation of a patient’s preferences, as an interpretation of personal value.

Broome’s suggestion cannot help one avoid our trilemma, however, as, given *Pluralism,* the trilemma arises also in cases where the same probability function is used when calculating the personal values of all patients. An analogous move to Broome’s, in the context of our trilemma, would be to argue that personal value must be such that it has an expectational form for all patients, i.e. to reject *Pluralism*.[[11]](#footnote-11) This possibility is discussed in the next section.

1. *Relaxing Pluralism*

As long as the personal values of all patients have an expectational form the trilemma does not arise (that is, there can be no cases in which *The Common Principle* conflicts with *ex-ante Pareto*).[[12]](#footnote-12) Thus, it is tempting to argue that *Pluralism* should be rejected.

Furthermore, there is a theoretical argument – often made in the context of utility and preferences, but one that can be extended to other interpretations of personal value – that supports this position. Here it is (formulated in terms of personal value): Suppose you rank all the prospects in some set of prospects according to how valuable they are for a given patient, and suppose this ranking obeys the axioms of orthodox decision theory. Then, it is possible to represent this ranking as a value ranking for a value function that does have an expectational form.[[13]](#footnote-13) Principles of distributive justice, such as *ex-ante Pareto* and *The Common Principle,* should be understood – so the argument goes – as referring to this type of representation of value-rankings (i.e. representations using value functions that have an expectational form). Since all value-rankings must respect the axioms of orthodox decision theory, it is always possible to use such representations and thus the trilemma can be avoided.

The standard reply to this argument (again, usually made in the context of utility and preferences, but readily extended to other interpretations of personal value) is that the axioms of orthodox decision- theory are, in fact, too demanding and, in some cases, violated by normatively permissible value rankings.[[14]](#footnote-14) In these cases it is impossible to represent these value rankings using value functions that have an expectational form.

I will not attempt to settle this debate here. However, I do believe that the case for *Pluralism* is stronger when it comes to many other interpretations of personal value (that is, other than utility). Specifically, this is so for any interpretation that is committed to, or at least supports, relying on a linear measure of personal value.

For example, many interpretations of personal value are committed to, or at least support, taking life expectancy as one of the indexes in a measure of personal value. If *Pluralism* is rejected, then it follows that, according to such interpretations, others things being equal, the prospect of living for 50 more years for certain is as valuable as the prospect of living for 100 more years with probability 0.5 and dying tomorrow with probability 0.5. This absurd conclusion can be avoided, of course, by arguing that the right measure of personal value is non-linear in years lived. This is the standard move made in the context of utility and the starting point for the debate regarding *Pluralism* in decision theory, mentioned above. However, part of the attraction of non-utility interpretations of personal value is that they point to concrete and easily measurable features of the world. Taking the non-linearity move undermines this advantage of such interpretations.

In any case, I believe that even those who reject *Pluralism* as a substantial assumption can benefit from using it as a methodological assumption in the context of our trilemma. Even those who insist that value-rankings must respect the axioms of orthodox decision theory, and thus can be represented using a value function with an expectational form, do not deny that such value-rankings can also be represented using value functions that do not have an expectational form. However, on the face of it, the normative statuses of both *The Common Principle* and *ex-ante Pareto* seem to be independent of the type of representation used for the patients’ value-rankings. Thus, by assuming *Pluralism* and then figuring out which one of the two other principles should be rejected when understood as applying to representations of the patients’ value-ranking that do not necessarily have an expectational form, one might better understand the normative statuses of these principles (and see the discussion in section 4).

1. *Relaxing ex-ante Pareto*

As mentioned above, while ex-post Pareto is relatively uncontroversial, the ex-ante principle is controversial. First, some (for example, Gilboa, Schmeidler and Samet 2004) argue (quite plausibly, I think) that it should be rejected in cases in which different patients have different beliefs (i.e. they assign different probabilities to the states) and replaced with a principle that is restricted only to cases in which all patients share the same probability function. This move cannot help us here, however, as the trilemma arises for the restricted ex-ante Pareto principle as well (as the example presented in section 1 shows). Mongin (2016) argues that even in some cases in which all patients use the same probability function but base their assignments of probabilities on different sets of evidence, the principle should be rejected. Although I think Mongin is right, the trilemma arises also in cases in which all patients have the same set of evidence and thus this move, too, cannot help us here.[[15]](#footnote-15)

Voorhoeve and Fleurbaey (2013) present a more general argument for rejecting *ex-ante Pareto*, which covers also cases in which all the patients share the same probability function and the same set of evidence. They argue that *ex-ante Pareto* should be rejected because it conflicts with ex-post egalitarianism (i.e. egalitarianism in terms of sure outcomes), even in cases in which the personal values of all patients have an expectational form. The example described in table 3 demonstrates such a case.

|  |  |  |
| --- | --- | --- |
|  | Probability 0.5 | Probability 0.5 |
| Distribution E | a, a | a, a |
| Distribution F | 2a+ɛ, 0 | 0, 2a+ɛ,  |

*Table 3*

It is easy to see that while the expected value of F (a + ɛ/2) is higher than that of E (a) for both patients, the choice between E and F is a choice between a certain {a,a} and a certain {2a+ɛ, 0}. However, according to ex-post egalitarianism, for at least some ɛ, the former distribution is more valuable than the latter. Thus, either ex-ante Pareto, or ex-post egalitarianism or the principle of state-wise dominance (that says that if one prospect is more valuable than another in every state, it is more valuable) must go. Voorhoeve and Fleurbaey argue that it is ex-ante Pareto that should be rejected.[[16]](#footnote-16)

Harsanyi (1955) famously used the same mathematical phenomenon to make the opposite move, i.e. to argue against ex-post egalitarianism based on its conflict with *ex-ante Pareto*. He also proved a more general result. He showed that in the case in which the distributer ranks different distributions in line with the axioms of orthodox decision theory, and the personal values of all the patients have an expectational form, the distributer’s ranking can be represented using a value function that is a weighted average of the patients’ personal values iff *ex-ante Pareto* is satisfied. Adding a simple symmetry axiom (that demands that all patients be given the same weight) lead him to accept Utilitarianism.[[17]](#footnote-17)

From this point of view, a possible way to understand the trilemma discussed here is as an argument in favor of Voorhoeve and Fleurbaey’s position. While Harsanyi relied on *ex-ante Pareto* in his justification of Utilitarianism, the trilemma shows that once we allow for pluralism regarding attitudes to risk, *ex-ante Pareto* conflicts not only with ex-post egalitarianism but also with Utilitarianism, which is Harsanyi’s own position. Thus, so the argument might go, it is *ex-ante* Pareto that must go, not ex-post egalitarianism.

While I think this line of argument is a successful one when considered in the context of the dialectic of the debate just described, it does not point to an explanation for why *ex-ante* *Pareto* is not a valid normative principle. The mere conflict with *The Common Principle* (or even only with ex-post egalitarianism) should, I believe, lead us to examine the normative basis of *ex-ante Pareto*, but in itself it is not a conclusive argument against it. What is needed, in other words, is an explanation for why, despite its strong intuitive appeal, the principle is invalid.

Furthermore, for those (such as Voorhoeve and Fleurbaey) who whish to keep ex-post Pareto while rejecting the ex-ante version, the explanation for the invalidity of the latter must be such that it does not apply to the former, i.e. it must point to the existence of uncertainty as the reason for *ex-ante Pareto’s* failure. Those who reject *ex-ante Pareto* only in cases in which different patients have different beliefs (or base their beliefs on different sets of evidence) have exactly this type of an explanation. They argue that, when it comes to distributive justice, mistaken beliefs or less than ideal sensitivity to the evidence should not matter. Thus, the distributer should not consider the patients’ personal values when calculated relative to their own beliefs, but rather the patients’ personal values when calculated relative to the “correct” probability function.[[18]](#footnote-18) However, this explanation is unavailable to those who reject *ex-ante Pareto* also in cases in which a single probability distribution is assumed and in which the set of evidence is shared among all patients (and the distributer).

Notice also that, when it comes to at least most of the central interpretations of “personal value”, ex-post Pareto is only an idealization. In real life, certain quantities of personal value are almost never distributed. What is in fact distributed are economic resources and, when those are understood as means of producing personal value, there is always some uncertainty involved. You receive some money – but you are uncertain about the inflation rate. You own a house – but you are uncertain whether it will burn down in a fire. You are chosen for a kidney-transplant operation – but you are uncertain whether (or when) your body will reject the kidney.[[19]](#footnote-19) Thus, rejecting *ex-ante Pareto* seems to amount to rejecting any practically-relevant version of the principle.

For this reason, I believe that *ex-post Pareto* and *ex-ante Pareto* rise and fall together. What, then, is the normative basis for the Pareto principle (in both its ex-post and ex-ante forms)? Two different types of justifications can be (and often – even if implicitly in some cases – are) given to the principle:consequentialist justifications and non-consequentialist ones.

Non-consequentialist justifications can take many forms. One might argue that the normative validity of the principle is rooted in respect for people’s preferences (see for example Broome 1987), or respect for people’s autonomy (Sugden 2004 can be interpreted along these lines). One might also argue that employing veil of ignorance reasoning must lead one to obey the principle, for if one is certain that – no matter which role in society one ends up occupying – one prospect, A, will be more valuable to one than another prospect, B, then A is more valuable for one than B even behind the veil of ignorance (i.e. by employing the veil of ignorance reasoning, *ex-ante Pareto* becomes an instance of the Dominance principle).

From the point of view of such non-consequentialist justifications for *ex-ante Pareto,* the trilemma can be understood as a demonstration of the way consequentialist considerations (as captured by *The Common Principle*, which seems to be a minimal consequentialist principle) can be in strong tension with non-consequentialist ones. This general lesson can be translated into specific arguments against some positions in the literature that employ (sometimes implicitly) both consequentialist and non-consequentialist elements in their justifications.

One such example is Buchak’s (2017) own attempt to argue for what she calls “relative prioritarianism” (roughly, the position according to which, when making distributional decisions, the interests of those who are relatively worse off should get higher weights than the interests of those who are relatively better off). Buchak shows that, given some plausible assumptions, veil of ignorance reasoning combined with the decision theory she proposed leads to relative prioritarianism. Buchak concentrates on cases in which no uncertainty regarding the state of the world, other than the uncertainty induced by the veil of ignorance, is involved. However, here we see that once uncertainty about the world is introduced into the model, Buchak’s theory together with a commitment to *ex-ante Pareto* (which, as explained above, seems to follow from the veil of ignorance assumption) must lead to the rejection of *The Common Principle*. However, *The Common Principle* is implied by relative prioritarianism. Thus, on the face of it, Buchak’s argument seems to be based on inconsistent assumptions.

To be fair to Buchak, the above objection does not threaten the version of the argument she actually presented. This is so in virtue of the specific form of the veil of ignorance assumption she employed. In Buchak’s argument, behind the veil of ignorance, a decision maker, not only does not have any knowledge of her own attitude to risk (i.e. in Buchak’s theory, of her own r(.) function), but also has no knowledge regarding the distribution of attitudes to risk (the r(.) functions) in the society. Buchak argues (quite convincingly, I think) that under such conditions a decision maker must adopt “the most risk avoidant attitude within reason” (see, for example, p. 18).

However, strikingly, the assumption regarding the lack of knowledge of the distribution of risk-attitudes in the society is disanalogous to the parallel assumption regarding utilities. When it comes to utilities, it is usually assumed (and must be assumed by Buchak for her argument to go through) that although the decision maker does not know her own utility behind the veil of ignorance, she does know the distribution of utilities in the society. Why, then, make a different assumption when it comes to risk-attitudes?

It seems to me that there is no reason to do so and this is so from the point of view of Buchak’s own justificatory strategy. The veil of ignorance assumption is usually understood as the theoretical vehicle through which the idea of impartiality is captured. However, in order to be impartial, it is enough to lack knowledge regarding one’s own relevant attitudes. Knowledge of the distribution of a given attitude (any attitude, not only attitudes to risk) in society does not make a decision maker any less impartial, as long as she does not know which position in society she occupies. Thus, the assumption of lack of knowledge regarding the distribution of attitudes to risk in society is unnecessary for expressing impartiality. Furthermore, Buchak’s own position is that people’s attitudes to risk should be respected when known. That is, when making choices for someone else, one should adopt the other persons’ attitude to risk, in case one knows it. Thus, it is clear that Buchak takes a person’s attitude to risk to be a morally significant feature of this person[[20]](#footnote-20) and, if this is so in a one-person case, it should be so also in cases that involve many people. Thus, there seems to be no reason to assume that the distribution of attitudes to risk in society is not known behind the veil of ignorance. As argued, in case it is known, *ex-ante Pareto* seems to follow and then relative prioritarianism, which is the position Buchak argues for, cannot be true.

Thus, it seems that arguing for prioritarianism (or any other theory that obeys *The Common Principle*) by employing the veil of ignorance reasoning cannot be a successful strategy for those who accept *Pluralism*.

Moving from an examination of the trilemma from the point of view of non-consequentialist justifications for the Pareto principle to an examination from a consequentialist point of view regarding Pareto leads, however, to more surprising conclusions. From a consequentialist point of view, it seems that the Pareto principle should be accepted as a principle of aggregation of personal value. The main idea is, I take it, that no feature of a distribution is valuable unless it is valuable to at least one of the patients (see, for example, Broome’s (1991 chapter 8) defense of what he calls “the principle of personal good”). While the Pareto principle does not logically follow from this idea, it does seem to be the most straightforward explication of it. To say that every valuable feature of a distribution is valuable for at least one patient seems to mean that the value of a distribution is increasing in the personal value of all patients, which is equivalent to the Pareto condition.

From a consequentialist point of view, then, the trilemma seems to be a puzzle, as *The Common Principle* too seems, on the face of it, to be a valid principle of aggregation of personal values. This apparent puzzle will be discussed in the next section.

1. *Relaxing The Common Principle*

I find it useful to think about *The Common Principle* in the context of our trilemma in terms of reasons for action, that is, taking the personal values of the patients to be representations of the strength of the reasons for action they have (or at least of the reasons for action they have that should be considered by a distributer). By adopting this perspective, one does not have to commit oneself to any specific position regarding the metaphysical priority relations between reasons and value. All that needs be assumed is that whenever one prospect is more valuable to a patient than another then this patient has stronger reasons to bring it about (if he could). This last assumption seems innocent.

Once adopting this point of view, a simple explanation for why, given *Pluralism*, *The Common Principle* is not a valid principle of aggregation of personal values suggests itself. As *The Common Principle* refers only to cases in which one distribution *necessarily* brings both a higher total sum of personal value than another distribution, and a more egalitarian distribution of this total sum, it is completely insensitive to the patients’ attitudes to risk. It is sensitive only to final distributions. However, given *Pluralism*, different patients can have different attitudes to risk. Now, when a patient is not risk-neutral (that is, when he assigns to an outcome of a prospect a weight which is different from its probability according to the prospect), it is, presumably, because he has some reason for this. Let us call such reasons “risk-related reasons”. Because *The Common Principle* is completely insensitive to the patients’ attitudes to risk, it fails to count these risk-related reasons. However, *Pluralism* allows for their existence. Thus, given *Pluralism, The Common Principle* fails as a principle of aggregation of personal values.

One can react to the above argument in two different ways. First, one can put the blame for the trilemma on *Pluralism*. One can understand the argument as an explanation for why *The Common Principle* is valid only when value-rankings are represented using value functions that have an expectational form. When the outcomes in a prospect are individuated in such a way that every reason for choice a patient has – including risk-related reasons – is included in the description of one of the prospect’s outcomes, the prospect’s personal value (for the patient) has an expectational form (because the strength of every risk-related reason is represented in the value assigned to one of the outcomes, and so representing it again by attributing to the patient some attitude to risk which is different from risk-neutrality, constitutes double counting).

A simple example might make the last point clearer. Consider again our original demonstration of the trilemma from the point of view of patient 2 (the risk-avoidant patient), only now we interpret the numbers in the table as representing years of life lived (which is, very plausibly, a constituent of personal value but not the only one). The two prospects (from the point of view of patient 2, and using – for convenience – a different scale) are represented in table 4.

|  |  |  |
| --- | --- | --- |
|  | State 1 (probability 0.5) | State 2 (probability 0.5) |
| A | 27 | 27 |
| B | 19 | 49 |

*Table 4*

As discussed in section 2, it seems perfectly reasonable (for, say, someone who is 50 years old) to value A (living for 27 more years for certain) more than B (living 19 more years with probability 0.5 and living 49 more years with probability 0.5). However, we can ask what it is about the two prospects that makes this valuation reasonable even though in B one’s life expectancy is higher. In other words, we can ask what are the risk-related reasons the patient has to choose A over B? There are many possible answers, of course, and different patients might have different risk related reasons. However, it seems that, given that we are able to point to these reasons, we can write them down in the description of at least one of the possible outcomes.

For example, suppose as patient 2 gets older he cares less about living every additional year of life and that this is a reason for him to value A more than B. In such a case, instead of identifying the possible outcomes with the numbers of years lived, we should identify the possible outcomes with some index that measures the value of the numbers of years lived when later years are discounted to the appropriate degree. Or suppose patient 2 disvalues being uncertain about when he will die because this makes him anxious. In such a case we should identify the possible outcomes with the number of years lived plus a reference to whether these years are lived with anxiety (e.g. the outcomes might be “27 years lived without anxiety”, “19 year lived with anxiety”, “49 years lived, 19 of which with anxiety”). Once we individuate the outcomes in such a way that all risk-related reasons are encoded in them, we are left with no reason to choose one of the prospects over the other over and above those that are represented by the values assigned to the possible outcomes. In such a case the patient’s value ranking must be represented using a value function that has an expectational form because there is no reason to attach to any of the outcomes a weight that is different from its probability.

Notice that according to this interpretation of the trilemma, although *Pluralism* is the principle that gets rejected, while *ex-ante Pareto* can be accepted without reservations, *The Common Principle* is accepted only when it is understood as referring to representations of value-rankings in terms of value functions that have an expectational form. For instance, if in our example the numbers in the table are understood as representing the strength of all relevant reasons *except risk-related reasons* the principle does not dictate valuing B more than A.[[21]](#footnote-21)

The argument just presented is based on two assumptions. The first is that it is always possible to individuate the outcomes in such a way that all risk-related reasons are included in the description of the outcomes. The second assumption is that once this is done, it is normatively impermissible for the patient to have any attitude to risk different from risk-neutrality. Both these assumptions can be questioned.

Regarding the first assumption, one might argue that there might be risk-related reasons that cannot be ascribed to any of the outcomes of a prospect because they are grounded in the risk itself, abstracted from the outcomes over which it is defined (see for example Bradley 2017). Furthermore, since (as argued above) when it comes to distributive justice, certainty is usually only an idealization, one should avoid giving the mere idea of “final outcomes” that do not involve any uncertainty, a substantial role in one’s normative arguments (otherwise one’s normative conclusions will apply only to an idealized world and not to the real one).

Regarding the second assumption, Murray and Buchak (2019) argue that different attitudes to risk are not always the result of different sets of risk-related reasons. Rather, in some cases two people who have exactly the same reasons (including risk related reasons) might have different attitudes to risk because they weigh these reasons in two different normatively permissible ways, under uncertainty.

Those who reject either one (or both) of these assumptions must conclude that *The Common Principle* is not a valid principle of aggregation of personal values (understood with respect to any representation of value-rankings). This, however, does not mean that they must reject *The Common Principle.* They might still accept it, but on other grounds.

The most natural way to go from here is to argue that *The Common Principle* is valid, not because it is sensitive in the right way to all of the patients’ personal reasons, but rather because it is sensitive to reasons that are not reasons of any patient in the group but are rather reasons the group has *as a group*. If we take both increasing the total sum of personal value and distributing it in a more egalitarian way as reasons *for the group* to prefer one distribution to another, then, trivially, the group has reasons to obey *The Common Principle.*

If this conclusion is the right one to draw then clearly it is the *Pareto* principle that should be rejected, but not only in its ex-ante but also in its ex-post form (but this, in light of the arguments in the previous section, is not surprising). It should be rejected because, by adopting the position according to which reasons that the group has as a group but that none of the patients has individually should be counted by a distributer, one commits oneself to a rejection of the idea that everything that is valuable in a distribution is valuable to at least one of the patients. Interestingly, while this is a commitment that some egalitarians (specifically those who believe that equality is valuable in itself, even when it is not valuable for any of the patients) are willing to take and for which they are often criticized by both utilitarians and prioritarians, here we see that once *Pluralism* is accepted, utilitarians and prioritarians must take it too.

The general lesson to be learnt from the discussion in this section, then, is that there is a close (and quite surprising) connection between the question of how to weigh reasons under uncertainty (and specifically whether or not there is a unique normatively permissible way to do so) and the question of whether groups’ interests (that are irreducible to individual’s interests) count for distributive justice considerations. For permisivists with respect to weighing reasons under uncertainty (i.e. those who believe that there is more than one permissible way to weigh reasons under uncertainty), it seems that the only (or at least, the most natural) way to accept *The Common Principle* (which, I believe, few philosophers would choose to reject) is to allow for groups’ intertest to count. Non-permisivists (who also believe that it is always possible to represent risk-related reasons in the description of one of the outcomes of a prospects) can accept a restricted version of *The Common Principle*, according to which the principle is valid only when understood as referring to those representations of the relevant value-rankings that have an expectational form.

1. *Conclusion*

Although I did not argue for any specific solution to the trilemma in this paper, I have reached several interesting conclusions.

First, by pointing to the trilemma I have demonstrated that the *ex-ante Pareto* principle is an even more demanding principle than is usually realized. It not only conflicts with ex-post egalitarianism and prioritarianism, but also, given *Pluralism,* with Utilitarianism. Thus, the trilemma can be used by opponents of the principle to strengthen their case against it.

Second, the trilemma shows that accepting the Pareto principle on non-consequentialist grounds is a radical move in the sense that it conflicts with minimal consequentialist considerations (captured by *The Common Principle).* I have shown how this general conclusion translates to a specific objection to Buchak’s argument for relative prioritarianism.

Finally, I have argued that accepting the Pareto principle on consequentialist grounds (i.e. accepting it as a valid principle of aggregation of personal values) has very strong implications. It commits one to being a non-primitivist regarding weighing reasons under uncertainty and to the position that it is always possible to individuate outcomes in such a way that every risk-related reason (in a given choice situation) is included in their descriptions.

Those who are unwilling to accept at least one of these two assumptions and are still committed to *The Common Principle* must reject the Pareto principle. I have argued that one tempting way to do this is to reject the idea that any valuable feature of a distribution is valuable for at least one of the patients.

*References*

Allais, M. (1953), Le Comportement de l’Homme Rationnel Devant le Risque: Critique des Postulats et Axiomesde l’Ecole Americaine, *Econometrica*, 21, 503–46.

Alon, S. and Gayer, G. (2016), Utilitarian preferences with multiple priors, *Econometrica,* 84 (3), 1181-1201,

Blackorby, C., D. Donaldson and P. Mongin (2004), Social Aggregation Without the Expected Utility Hypothesis, *Cahier no. 2004-020*, Laboratoire d'économétrie, Ecole Polytechnique.

Bradley, R. (2005), Bayesian Utilitarianism and Probability Homogeneity, *Social Choice and Welfare*, 24, 221-51.

Bradley, R. (2017), *Decisions Theory with a Human Face*, Cambridge University Press.

Briggs, R. (2015), Cost of abandoning the Sure-Thing Principle, *Canadian Journal of Philosophy,* 45 (5-6), 827-40.

Broome, J. (1987), Utilitarianism and Expected Utility, *The Journal of Philosophy,* 84 (8), 405-22.

Broome, J. (1990), Bolker–Jeffrey Expected Utility Theory and Axiomatic Utilitarianism*, Review of Economic Studies*, 57, 477–03.

Broome, J. (1991), *Weighing Goods*, Cambridge, Mass.: Basil Blackwell.

Buchak, L. (2013), *Risk and Rationality*, Oxford University Press.

Buchak, L. (2015), Revisiting Risk and Rationality: A Reply to Pettigrew and Briggs, *Canadian Journal of Philosophy,* 45 (5-6), 841-62.

Buchak, L. (2017), Taking risks behind the veil of ignorance, *Ethics,* 127 (3), 610-44.

Ellsberg, D. (1961), Risk, Ambiguity, and the Savage Axioms, *Quarterly Journal of Economics*, 75, 643–69.

Fleurbaey, M. (2010), Assessing Risky Social Situations, *Journal of Political Economics,* 118 (4), 649-78.

Fleurbaey, M. and Mongin, P. (2016), The Utilitarian relevance of the aggregation theorem, *American Economic Journal: Microeconomics,* 8 (3), 289-306.

Fleurbaey, M. and A. Voorhoeve. (2013), Decide as you would with full information! An argument against ex ante Pareto. In *Inequalities in Health: Concepts, Measures and Ethics*, ed. N. Eyal, S. Hurst, O. Norheim and D. Wikler, 113–28, Oxford University Press.

Frick, J. (2015), Contractualism and Social Risk, *Philosophy and Public Affairs,* 43 (3). 175–223.

Gilboa, I., D. Schmeidler and D. Samet (2004), Utilitarian Aggregation of Beliefs and Tastes, *Journal of Political Economy*, 112, 932-38.

Goldschmidt, Z. and Nissan-Rozen, I. (2020), The Intrinsic Value of Risky Prospects, *Synthese,* https://doi.org/10.1007/s11229-020-02532-3.

Harsanyi, J.C. (1955), Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility, *Journal of Political Economy*, 63, 309-21.

Jeffrey, R. (1965), *The Logic of Decision*, University of Chicago Press.

Kopec, M. and Titelbaum, G. M. (2016), The Uniqueness Thesis, *Philosophy Compass,*11 (4), 189-200.

Machina, M.J. (2008), Non-expected Utility Theory, in Palgrave Macmillan (eds) *The New Palgrave Dictionary of Economics,* Palgrave Macmillan, London.

Mongin, P. (1995), Consistent Bayesian Aggregation, *Journal of Economic Theory*, 66, 313-51.

Mongin, P. (2016), Spurious Unanimity and the Pareto Principle, *Economics and Philosophy,* 32 (3), 511-32.

Mongin, P. and Pivato, M. (2015), Social Preferences and Social Welfare Under Risk and Uncertainty, in Adler, M. and Fleurbaey, M. eds. *Oxford Handbook of Well-Being and Public Policy,* Oxford University Press.

Murray, D. & Buchak, L. (2019), Risk and Motivation: When the Will is Required to Determined What to Do, *Philosophers’ Imprint,* 19 (16), 1-16.

von Neumann, J. and Morgenstern, O. (1947), *Theory of Games and Economic Behavior*, (second edition), Princeton University Press.

Nissan-Rozen, I. (2017), How to Be an Ex-Post Egalitarian and an Ex-Ante Paretian, *Analysis*, 77 (3), 550–58.

Parfit, D. (1984), *Reasons and Persons,* Oxford University Press.

Parfit, D. (1997), Equality and Priority, *Ratio,* 10 (3), 202-21.

Pettigrew, R. (2015), Risk, Rationality and Expected Utility Theory, *Canadian Journal of Philosophy,* 45 (5-6), 798-826.

Rabin, M. (2000), Risk Aversion and Expected Utility Theory: A Calibration Theorem, *Econometrica*, 68, 1281–92.

Rowe, T. (2019), Risk and the unfairness of some being better off at the expense of others, *Journal of Ethics and Social Philosophy,* 16 (1), 44-66.

Savage, J. L. (1954), *The Foundations of Statistics*, Dover.

Stefánsson, H.O. and Bradley, R. (2015), How valuable are chances?, *Philosophy of Science*, 82 (4), 602-25.

Stefánsson, H.O. & Bradley, R. (2017), What is Risk Aversion? *The British Journal for the Philosophy of Science,* 70 (1), 77-102.

Stefánsson, H.O. (2018), Is risk aversion irrational? Examining the “fallacy” of large numbers, *Synthese.* https://doi.org/10.1007/s11229-018-01929-5.

Sugden, R. (2004), The Opportunity Criterion: consumer sovereignty without the assumption of coherent preferences, *American Economics Review,* 94 (4), 1014-1033.

Thoma, J. (2019), Risk aversion and the long run, *Ethics,* 129 (2), 230-53.

Thoma, J. and Weisberg, J. (2019), No escape from Allias: Reply to Buchak, *Philosophical Studies.* https://doi.org/10.1007/s11098-019-01322-z.

1. \* Earlier versions of this paper were presented at the Center for Values and Social Policy, the University of Colorado, Boulder and at the Normative Ethics Business Workshop Series of the Carol and Lawrence Zicklin Center for Business Ethics Research. I thank the participants of these events, as well as Neta Bar-Ziv, Jonathan Fiat and two anonymous referees for their helpful comments. My research was supported by the Israeli Science Foundation, Grant Number 327/18. [↑](#footnote-ref-1)
2. I deliberately present the trilemma in an informal way. It can be formalized in different ways, but each unavoidably involves relying on some structural technical assumptions that would have to be explained at some length. As the aim of this paper is to explore the trilemma from a normative point of view, I chose not to commit myself to any specific formalization. The price of this choice is that, in some parts, the discussion involves some imprecise claims and uses some vague concepts. I explicitly highlight those places in the discussion where I believe this potentially hides some substantive normative issue. Throughout this paper I assume that personal value is cardinally measurable and interpersonally comparable and that the uncertainty (or risk) involved can be represented using a single probability distribution over a set of states. Although I did not find any formulation of this specific trilemma in the literature, the interested reader can find related results in the economic literature. For an excellent review see Mongin and Pivato 2015. [↑](#footnote-ref-2)
3. The term “in a more egalitarian way” is ambiguous, of course, but (as demonstrated below) the trilemma arises even in cases in which it is completely uncontroversial that the distribution with the higher total sum of personal value is also the more egalitarian one. Thus, there is no need for me to commit myself to a specific measure of inequality. [↑](#footnote-ref-3)
4. Rejecting the restrictive part of the principle (the one that demands that the value of a prospect is a weighted sum of the values of the prospects’ outcomes) cannot help one avoid the trilemma (and, in any case, as an anonymous referee commented, rejecting this part amounts to rejecting state-wise dominance on the level of personal value, which is highly unintuitive). The problematic part is the permissive one (the one that allows the weights to be different from the probabilities). Below I show that it is possible to build a distribution problem in which the conflict between *The Common Principle* and *ex-ante Pareto* arises in every case in which one of the patients is risk neutral (i.e. the value of a prospects for her is the prospect’s expected value) and the other patient assign to the states weights which are different from their probabilities, *for any weights different from the probabilities.* I do not even have to assume that the weights sum to 1. The trilemma, however, does not arise when the weight each outcome gets is equal to its probability. This will be discussed in section 2. The probability in question can be understood either as objective probability or as the subjective probability of the distributer or in any other way according to which the same probability function is used when assessing the prospects from the point of view of all patients. I discuss this below. [↑](#footnote-ref-4)
5. I thank an anonymous referee for pointing this out to me. [↑](#footnote-ref-5)
6. At least some of the credit for this should be given to Parfit’s (1997) famous “levelling down objection”. [↑](#footnote-ref-6)
7. That is, it is very hard to explain several common patterns of choice, such as, for example, the ones described by Allias (1953) Elsberg (1961) and Rabin (2000), while holding it. See Machina (2008) for a good review. [↑](#footnote-ref-7)
8. In fact, in Buchak’s theory (and in rank-dependent decision theories in general) the weights are not attached to the states, but rather to value levels, i.e. to the different possible levels of value the patient might gain from the prospect. However, when it comes to the simple example below and to its generalization (that follows) *Pluralism* captures these theories too. Since any plausible decision theory should cover such simple examples, the differences between rank-dependent decision theories and the theories characterized by *Pluralism* should not concern us. [↑](#footnote-ref-8)
9. And notice that had patient 2 used r(p) = p (i.e. had he been neutral to risk), Buchak’s theory would reduce to orthodox decision theory as 1.9 + 0.5(3) = 1.9 + 0.5(4.9 -1.9) = 0.5\*1.9 + 0.5\*4.9. This is generally true in Buchak’s theory. [↑](#footnote-ref-9)
10. For formal discussions of related results see Broome (1990), Mongin (1995) and Bradley (2005), among others. [↑](#footnote-ref-10)
11. Alternatively, one can demand that, while the personal values of all patients might have a non-expectational form, this form must be identical for all patients. Interestingly, while taking this route cannot help one avoid the trilemma in cases in which all patients are risk-seeking, it might work when all patients are risk-avoidant (but this seems to depends on the formal framework used). Here is an example for the risk-seeking case.

|  |  |  |
| --- | --- | --- |
|  | 0.5 | 0.5 |
| K | 5, 14 | 14,5 |
| L | 10,10 | 10,10 |

In case both patients obey Buchak’s theory and use r(p) = p0.5 the personal value from K is higher than that of L for both of them, but L brings a higher total sum of personal value, which is distributed in a more egalitarian way in both states.

The risk-avoidance case is non-trivial. When it comes to Buchak’s framework, when all patients use the same r(.) function and this function is convex, the trilemma does not arise. This is proven and some of its normative implications are discussed in a companion paper (under review). In any case, when it comes to the trilemma discussed here, demanding both that the functional forms of the personal values of all patients will be identical and that it will be a risk-avoidant one seems hard to justify as it seems at least permissible for a patient to be risk-neutral (i.e. to obey orthodox decision-theory). [↑](#footnote-ref-11)
12. This is a consequence of Harsanyi’s (1955) Utilitarian Theorem. [↑](#footnote-ref-12)
13. This is what decision theory’s representation theorems say. There are many such theorems, made in different frameworks, that use different sets of axioms. See for example, von Neumann and Morgenstern (1947), Savage (1954) and Jeffrey (1965). [↑](#footnote-ref-13)
14. For some discussion see Buchak (2013) and (2015), Stefansson and Bradley (2015) and (2017), Briggs (2015), Pettigrew (2015), Stefansson (2018), Thoma (2019), Thoma and Weisberg (2019), and Goldschmidt and Nissan-Rozen (2020). [↑](#footnote-ref-14)
15. More generally, to accept *Pluralism* (which is a normative principle) is to accept that even when there is no disagreement on the epistemic level between different patients, they can assign different weights to the states. This blocks (for those who accept *Pluralism*)the general line of objection to *ex-ante Pareto* (suggested by Mongin 2016) according to which the principle is invalid when the agreement on the epistemic level is a result of deeper disagreements that cancel each other out. I thank an anonymous referee for pushing me on this point. [↑](#footnote-ref-15)
16. For some discussions see Frick (2015), Nissan-Rozen (2017) and Rowe (2019). The first two argue – in different ways – that it is state-wise dominance that should be relaxed (in the relevant class of cases). [↑](#footnote-ref-16)
17. Many economists have presented generalizations of Harsanyi’s result and explored different ways of relaxing some of Harsanyi’s assumptions. See for example, Fleurbaey (2010), Fleurbaey and Mongin (2016), Alon and Gayer (2016). [↑](#footnote-ref-17)
18. Which probability distribution is the correct one might, of course, be a matter of dispute in many cases, but the argument only needs the assumption that such a single normatively correct probability distribution exists in the relevant class of cases. Thus, philosophers who rejects “the uniqueness thesis” (see Kopec and Titelbaum 2016 for a review of the literature) might hesitate to follow this line of reasoning. [↑](#footnote-ref-18)
19. See Broome (1990) and Mongin (2016) for a similar point. [↑](#footnote-ref-19)
20. Indeed, this seems like a very natural assumption to make, at least in light of Buchak’s decision theory. However, see the discussion in the next section. [↑](#footnote-ref-20)
21. An anonymous referee pointed out that the same type of maneuver just discussed with respect to *The Common Principle* can be made also with respect to *ex-ante Pareto,* i.e. to accept *The Common Principle* without reservations, but *ex-ante Pareto* only when understood as referring to representations of value-rankings in terms of value functions that have an expectational form. I agree with the referee that this possibility naturally suggests itself, but I do not find it very promising. This is so because, unlike *The Common Principle, ex-ante Pareto* aggregates values across patients; not across states and so its validity seems to be independent of the representation used. The reason the maneuver seems promising in the case of *The Common Principle* is that while under the expected value representation, the strength of all reasons, including risk-related reasons, is represented by the values attached to the different outcomes, under non-expected value representations this is not the case. Under such representations some risk-related reasons are represented by the way the weights are assigned to the different states. Thus, when aggregating across states the type of representation used matters. However, no analogous move is available when it comes to *ex-ante Pareto*. The value of a prospect for a patient is independent of the way an outside observer represents it. I thank the referee for pressing me on this point. [↑](#footnote-ref-21)