

## ON *EX ANTE* CONTRACTUALISM

*Korbinian Rüger*

CONTRACTUALISM is a claims-based model of moral rightness. It is the view, brought forward most notably by T.M. Scanlon, that an action is right if and only if it is justifiable to all. An action is justifiable to all just when it is licensed by a principle that cannot be reasonably rejected by any single individual.<sup>1</sup> Further, a principle can only be reasonably rejected for personal reasons. Contractualism thus construed excludes impersonal reasons derived from, for example, the overall value of an outcome. It thereby denies the permissibility of interpersonal aggregation of harms and benefits to determine which action is right. In situations in which individuals have competing claims to be helped, we always ought to pursue the policy that satisfies the single strongest individual claim, or, in converse, minimizes the strongest individual complaint against it, by following “the principle whose implications are most acceptable to the person to whom it is least acceptable.”<sup>2</sup>

This implication of contractualism clearly demarcates the view from thoroughly aggregative theories like utilitarianism. I here understand utilitarianism as standard act utilitarianism, where we always ought to pursue the action that will lead to the greatest (expected) sum total of well-being. The difference between the two rival theories becomes apparent in cases like:

*Death versus Headaches:* We can either save Ann from a terminal illness or prevent any number of different people from suffering a mild headache.

By virtue of what Ann stands to lose, her claim to be saved from death is clearly greater than any other individual claim to be spared a headache. Under contractualism we therefore ought to save her. This is the case irrespective of how many people stand to suffer a headache. Under utilitarianism, on the other hand, our answer will depend on the number of people that we could spare the headache. For some number of people, the benefits derived from the spared headache will *in sum* outweigh the benefit to Ann if we choose to save her. Contractualism

1 See Scanlon, *What We Owe to Each Other*, 189–248.

2 Kumar, “Risking and Wronging,” 31.

demands what many people take to be the obviously correct choice in *Death versus Headaches*.<sup>3</sup>

This is straightforward in hypothetical situations of absolute certainty like the above. The approach, however, is less clear about situations in which we do not yet know the outcomes our choices will lead to. These cases, however, are much more common. With Barbara Fried, one could even say that

in the real world, no conduct, judged *ex ante*, is certain to harm others. This is true even of harms that are intended. . . . If I point a gun at your head and pull the trigger, I am overwhelmingly likely to kill or seriously injure you, but I am not certain to do so. The gun could misfire, I could have forgotten to load it, [etc.].<sup>4</sup>

So rather than occurring with certainty, most harms result from risks that have been imposed on people or have not been eliminated. It is therefore imperative for contractualists to offer an account of how their theory deals with risk.

Johann Frick has developed such an account: *ex ante* contractualism.<sup>5</sup> In brief, *ex ante* contractualism holds that in situations involving risk we ought to act in accordance with principles that license the action that satisfies the strongest individual claim, where those claims are a function of the expected value that a given policy gives each person *ex ante*. It thus offers an alternative to the *ex post* reasoning employed by other contractualists, most notably Scanlon himself.<sup>6</sup>

I here challenge Frick's version of *ex ante* contractualism on contractualist grounds.<sup>7</sup> My argument proceeds as follows. In the first section I distinguish between *ex ante* and *ex post* contractualism in more detail. In the second section I argue that adopting *ex ante* contractualism would have far-reaching implications that contractualists would find very hard to accept. I show that *ex ante* contractualism in fact includes an implicit appeal to the interpersonal aggregation of

3 See, e.g., Voorhoeve, "Why One Should Count Only Claims with Which One Can Sympathize."

4 Fried, "Can Contractualism Save Us from Aggregation?" 50.

5 Frick, "Contractualism and Social Risk." See also Frick, "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives." Unless noted otherwise, henceforth "*ex ante* contractualism" refers to Frick's version of the view.

6 See Scanlon, *What We Owe to Each Other*, 189–248. See also Reibetanz-Moreau, "Contractualism and Aggregation." Note that Scanlon has since changed his position, crediting an earlier version of Frick's article. See Scanlon, "Reply to Zofia Stemplowska." For a critique of the *ex post* approach, see Ashford, "The Demandingness of Scanlon's Contractualism."

7 My argument is specifically addressed at Frick's way of developing *ex ante* contractualism. It is possible that there is another way of developing the view to which my argument does not apply. I do not pursue this possibility here.

harms and benefits. In the third section I show that Frick's argument for the principled priority of identified over unidentified lives, another troubling implication of *ex ante* contractualism, is unsound. In the fourth and final section I briefly comment on a possible pluralistic approach to get around some of the defects of *ex ante* contractualism. I conclude that, to deal with uncertainty, contractualists should not adopt *ex ante* contractualism, at least not Frick's version. Rather, they should adopt a suitably amended *ex post* approach.

### 1. EX ANTE AND EX POST CONTRACTUALISM

Let me introduce the *ex post* approach and then contrast it with the *ex ante* approach by way of one of Frick's examples.

*Mass Vaccination:* One million children are threatened by a virus, which will kill all of them if we do nothing. We must choose between producing one of three vaccines:

- Vaccine 1 is certain to save every child's life. However, if a child receives Vaccine 1, the virus will permanently paralyze one of the child's legs.
- Vaccine 2 gives every child a 99.9 percent chance of surviving the virus completely unharmed. However, for every child there is a corresponding 0.1 percent chance that Vaccine 2 will be completely ineffective. (Assume that the outcomes for different children are probabilistically independent.) Call the children who end up dying the *luckless children*.
- Vaccine 3 is sure to allow 999,000 children to survive the virus completely unharmed. However, because of a known particularity in their genotype, Vaccine 3 is certain to be completely ineffective for 1,000 identified *doomed children*.<sup>8</sup>

First consider a choice between only Vaccines 1 and 3 ( $V_1$  and  $V_3$ ). Here we are not dealing with uncertainty and it is straightforward what contractualism recommends. If we choose  $V_1$ , no single child will have a complaint that is as strong as the individual complaints of the doomed children if we choose  $V_3$ .<sup>9</sup>

8 See Frick, "Contractualism and Social Risk," 181–83. Note that Frick presents two distinct cases, in both of which Vaccine 1 is available, but Vaccines 2 and 3 only are available in one case.

9 Frick writes: "The individual burden of becoming paralyzed in one leg, though significant, is not even close to that of losing one's life at a young age" ("Contractualism and Social Risk," 183). Note that this information underspecifies (or even ill specifies) the strength of the individual complaints. If we assume a counterfactual account of harm, the complaints of the doomed children if we pick  $V_3$  over  $V_1$  are not complaints against being left to die, where

We therefore ought to choose  $V_1$ . If, on the other hand, we consider a choice between  $V_1$  and  $V_2$ , we are entering the territory of risk and things are less clear. This is because there are two ways of singling out the relevant complaints that we should take into account. Under one interpretation we look at the outcome that a given vaccine will produce and look at the single strongest complaint any individual will have in that outcome. If we choose  $V_2$ , we expect one thousand children to die.<sup>10</sup> Though we do not know how many children exactly will die, it is statistically certain that at least one child will die.<sup>11</sup> Since we are concerned with the single strongest individual complaint, this is all we need to know. Like in  $V_3$ , this complaint will be stronger than any complaint under  $V_1$ . Again, we ought to choose  $V_1$ . This is the *ex post* approach.

According to the *ex ante* approach, on the other hand, the relevant complaints are a function of the expected value an action gives each individual before it is performed. Under this account, a complaint against being subjected to a risk of suffering a harm is the complaint against being subjected to that harm with certainty discounted by the unlikelihood of the harm actually occurring. In *Mass Vaccination* the individual *ex ante* complaints against  $V_2$  are thus only 0.1 percent as strong as a complaint against dying from the virus with certainty. The strongest *ex ante* complaint against  $V_2$  is therefore much smaller than the strongest *ex ante* complaint against  $V_1$ , which in turn is smaller than the strongest *ex ante* complaint against  $V_3$ .

Accordingly, Frick's account selects  $V_2$  over  $V_1$ ,  $V_1$  over  $V_3$ , and  $V_2$  over  $V_3$ . This ensures that in each choice we minimize the strongest *ex ante* complaint. The *ex post* approach on the other hand would choose  $V_1$  over  $V_2$  and  $V_3$ , and would likely be indifferent between  $V_2$  and  $V_3$ .

*Mass Vaccination* thus shows how the *ex ante* and *ex post* approaches come apart. According to Frick it also shows why *ex post* contractualism is unattractive. It fails to make a principled distinction between  $V_2$  and  $V_3$ . As long as we know that *someone* will die if we pick  $V_2$ , and therefore they have a stronger complaint than anyone else if we choose  $V_1$ , this is enough for *ex post* contractualism to rule out  $V_2$ . It fails to take into account the special predicament the doomed children find themselves in under  $V_3$  as it assimilates their fate to those of the luckless

---

the alternative would be life in full health, but complaints against being left to die, where the alternative would be life with one paralyzed leg. Such complaints are presumably much weaker. I think this point is overlooked by Frick. Nonetheless it is reasonable to assume that these weaker complaints are still decisively stronger than the complaints of the other (non-doomed) children against a policy that leaves them with one paralyzed leg, where the alternative would be life in full health.

10  $0.001 \times 1,000,000$ .

11  $1 - (999/1,000)^{1,000,000}$ .

children in  $V_2$ . Frick would say that it fails to distinguish between the fact that “we know that someone will die” ( $V_2$ ) and the fact that “there is someone whom we know will die” ( $V_3$ ).<sup>12</sup>

Because of these alleged shortcomings of *ex post* contractualism, Frick proposes his *ex ante* approach. The main argument for this approach is the *argument from the single-person case*.<sup>13</sup> According to this argument, if we have an option available that is in the best *ex ante* interest of all individuals, we ought to choose it. We ought to adhere to the *ex ante* Pareto principle.

*Ex Ante Pareto Principle:* If an alternative has higher expected utility for every person than every other alternative, then this alternative should be chosen.<sup>14</sup>

Frick argues that the argument from the single-person case establishes the *ex ante* Pareto principle as a principle of contractualist ethics. We can decompose cases like *Mass Vaccination* into a large number of single-person gambles. Suppose again that we are facing a choice between  $V_1$  and  $V_2$  (recall that *ex post* contractualism chooses  $V_1$ ). This choice can be broken down into one million single-person cases. Suppose that Ann is one of the affected children and we ask ourselves what we would choose if we were solely motivated by her self-interest. We know that  $V_1$  will let her survive the virus but leave her with one paralyzed leg and that  $V_2$  will let her survive the virus completely unharmed with probability 999/1,000 and will lead to her death with probability 1/1,000. Given reasonable assumptions about which level of well-being (or utility) these three possible outcomes would deliver, we can calculate the expected value of both options. Suppose we assume that, for Ann, life with one paralyzed leg is four-fifths as good as life at full health, which we can arbitrarily fix to utility level 10, with death corresponding to 0. The expected utility of  $V_1$  then is 8, while the expected utility of  $V_2$  is 9.99.<sup>15</sup> Thus, the expected utility of  $V_2$  for Ann exceeds that

12 See Frick, “Contractualism and Social Risk,” 200, and “Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives,” 193. I do not take this distinction to be morally as important as Frick thinks it is. I shall not argue for this claim directly, though. Rather I will show that there are cases where even Frick’s own account fails to make the distinction.

13 See Frick, “Contractualism and Social Risk,” 186–94, and “Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives,” 133. For similar arguments see Dougherty, “Aggregation, Beneficence, and Chance,” and Hare, “Should We Wish Well to All?”

14 This formulation is taken from Fleurbaey and Voorhoeve, “Decide as You Would with Full Information!” 114.

15  $\frac{4}{5} \times 10$  and  $\frac{999}{1,000} \times 10 + \frac{1}{1,000} \times 0$ .

of  $V_1$ , and, if we are only concerned with her best interest, we ought to choose  $V_2$ .<sup>16</sup> This seems to be the right course of action. After all, what other than Ann's best interest would we base our decision on?

But, of course, this reasoning is correct for every single child in *Mass Vaccination*, where the possible outcomes and corresponding odds are exactly the same as in the one-person case. Thus, if we are concerned with every child's best interest, we ought to choose  $V_2$ , just like we ought to choose  $V_2$  in the one-person case when we are only concerned with Ann's best interest. The contractualist rationale behind this is that choosing  $V_2$  is the only action that is justifiable to all. Whatever the outcome of choosing  $V_2$ , we can offer each child the following justification: "When we had to choose, we did what was in your own best interest." This justification is not available to us if we choose  $V_1$ . I confess that I find this argument very seductive. In the following I argue, however, that contractualists ought to reject it and with it the *ex ante* Pareto principle.

## 2. THE IMPLICATIONS OF EX ANTE CONTRACTUALISM

Return to *Mass Vaccination*. Only now suppose that instead of  $V_3$ , we have  $V_3^*$  available. Like  $V_3$ ,  $V_3^*$  is sure to allow 999,000 children to survive the virus completely unharmed. However, because of a certain particularity in their genotype,  $V_3^*$  is certain to be completely ineffective for one thousand *unidentified* doomed children, instead of *identified* doomed children. We can imagine, for example, that we have tested all one million children for that genotype and have found that the vaccine will be ineffective for exactly one thousand of them. However, before we communicated the test results to anyone, our system broke down and we now have no way of assigning the positive results to any particular children.

Given that we chose  $V_2$  over  $V_3$ , should we now choose  $V_2$  or  $V_3^*$ ? In order to answer that question we need to investigate whether  $V_3^*$  is relevantly different from  $V_3$ . Only if it is can we justify choosing  $V_2$  over  $V_3$ , being indifferent between  $V_2$  and  $V_3^*$ . If  $V_3^*$  is not relevantly different, then, given that we chose  $V_2$  over  $V_3$ , we also ought to choose  $V_2$  over  $V_3^*$ . In this section I will argue that, first, we should not judge  $V_3$  and  $V_3^*$  differently; second, that *ex ante* contractualism, however, is committed to doing so; and, third, that this puts the account in a precarious position.

To me,  $V_3^*$  seems like  $V_3$  in all important respects. In  $V_3^*$  as in  $V_3$ , we know the exact outcome. We know that exactly one thousand children are going to die and that for them the vaccine was always going to be ineffective. Like with

16 Note that this result will be achieved even if Ann considers life with one paralyzed leg only slightly worse than life at full health.

$V_3$ , these children are doomed to die if we choose  $V_3^*$ . I therefore cannot see why we should choose  $V_2$  over  $V_3$ , but be indifferent between  $V_2$  and  $V_3^*$ , i.e., prefer to have  $V_3^*$  rather than  $V_3$  available. To see that this is implausible, suppose that we not only have  $V_3^*$  but two different vaccines— $V_3^*$  and  $V_3^{**}$ —available. However, for each of these vaccines it will (very likely) be an entirely different group of one thousand children for whom the vaccine will not work and who will be killed by the virus. Obviously we have no reason to choose any one of these vaccines over the other. Whatever we do, one thousand unknown, doomed children are going to die. The vaccines are equally choice worthy and we should randomize.

Suppose that we settle on  $V_3^*$ . Before we actually administer the vaccine, however, we learn who the children are for whom  $V_3^*$  will not do anything (maybe we were able to restore our database for  $V_3^*$ ). Should we now because of that switch to  $V_3^{**}$ ? I think clearly not. This would be an unnecessary “second lottery” and would arbitrarily favor those children for whom  $V_3^*$  is ineffective to the disadvantage of those children for whom  $V_3^{**}$  is ineffective. Nothing about the vaccines has changed and we said above that they are equally choice worthy. They still are. This, however, is in effect the same situation we face when comparing Frick’s  $V_3$  and my  $V_3^*$ . We therefore ought not to judge  $V_3$  and  $V_3^*$  differently.

Frick’s account, however, is committed to judging  $V_3$  and  $V_3^*$  differently. It is committed to judging  $V_3$  impermissible, but  $V_3^*$  (along with  $V_2$ ) permissible. This is because the argument from the single-person case applies to  $V_3^*$  as it applies to  $V_2$ . Here too it would be in each individual child’s best interest to choose the risky vaccine over  $V_1$ . The expected value of  $V_3^*$  for each individual child is the same as  $V_2$ .<sup>17</sup>

Though Frick does not consider  $V_3^*$ , he considers a nearby case. This case is like my  $V_3^*$ , only here there is a test we could carry out to identify the doomed children, but it would be very expensive. Frick argues that in this case adminis-

17 *Ex ante* contractualists could reply that there is one important difference between  $V_2$  and  $V_3^*$  that I have overlooked—namely, that while in  $V_3^*$  it is merely *epistemically* uncertain who will die, in  $V_2$  it is *objectively* (or *physically*) uncertain who will die. Frick, however, carries out his discussion on the assumption that *all* probabilities are merely epistemic. He writes that “when using the terms ‘probability’ or ‘chance’ ... I assume that we are speaking not about objective indeterminacy at the level of physical reality itself, but about epistemic probability” (Frick, “Contractualism and Social Risk,” 182). He furthermore argues, rightly I think, that for the moral assessment of risky policies this distinction makes no difference. (See Frick, “Contractualism and Social Risk,” 197–201.) In any case, it is doubtful whether objective probabilities at the physical level even exist. (See, e.g., Lewis, “A Subjectivist’s Guide to Objective Chance.”) Letting one’s moral theory depend on the assumption that they do exist seriously diminishes its attractiveness.

tering the vaccine would be justifiable to all and therefore permissible. In such a case we can say to each child “given justifiable limits on the resources we can be expected to expend in gathering further information about your particular case, [the vaccine] is highly likely to benefit you, and has only a tiny chance of turning out to your disadvantage.”<sup>18</sup> If administering the vaccine where it is *very costly* to find out which children will not be helped by it is permissible, then *a fortiori* administering  $V_3^*$ , where it is *impossible* to find out which children carry the problematic gene, must also be permissible. I now argue that this judgment concerning  $V_3^*$  spells trouble for the account. Consider the following case.

*Glass Box Villain (Known Victim)*: An evil villain has taken twenty-six hostages named Ann, Bob, Carl . . . and Zeta. He places you in the following diabolic choice situation: he has placed all of them in twenty-six individual glass boxes standing up side by side. The last box is made of regular glass and the other twenty-five boxes are made of extra-heavy glass. You can see that Zeta is placed in the last box. The villain asks you to decide between the following two options: (1) he will either fire a shot at her box or (2) fire twenty-five individual shots at the other boxes. If he fires at Zeta’s box, the bullet will not be stopped and Zeta will be killed. If he fires at the twenty-five boxes made of extra-heavy glass, the glass will divert the bullets. However, the glass will crack and the debris will disfigure the twenty-five hostages in a way that permanently leaves them at a well-being level 9.5 on a scale from 0 to 10, where 10 corresponds to a life in full health and 0 to death. If you refuse to decide, the villain will blow up all boxes, killing all twenty-six hostages. How should you decide?

I assume that refusing to decide should be ruled out as an option. Between the two remaining options, it is clear what contractualism tells you to do. You should choose (2). Choosing (1) would kill Zeta only to save twenty-five other people from a relatively minor harm. The complaints of the twenty-five on you are not even close to Zeta’s complaint. And since contractualism prohibits you from aggregating the twenty-five weak complaints to outweigh Zeta’s strong complaint, you ought to save Zeta’s life and let the villain fire at the twenty-five boxes made of extra-heavy glass. Since there is no uncertainty involved, *ex ante* and *ex post* contractualism do not come apart in this case. Consider, however, the following variation of the case.

*Glass Box Villain (Unknown Victim)*: Everything is as before, only now the boxes are opaque and neither you nor the hostages know whether it is

18 Frick, “Contractualism and Social Risk,” 194.



Zeta or any of the other twenty-five in the box made of regular glass. How should you decide?

I think if in *Glass Box Villain (Known Victim)* you ought to stop the villain from firing at the last box, then in this case you ought to act in the same way. I cannot possibly see why the fact that the twenty-six boxes are now opaque should change our moral assessment of the case in any way. (Remember that there is nothing about Zeta as a person that should make us favor her over the other twenty-five hostages in any way.) However, Frick’s *ex ante* contractualism is committed to the view that while in *Glass Box Villain (Known Victim)* you ought save Zeta, in *Glass Box Villain (Unknown Victim)* you ought to let the villain kill the person in the last box.

It is so committed because in *Glass Box Villain (Unknown Victim)* for all you know it could be any of the twenty-six hostages in the last box. In this respect it is parallel to  $V_3^*$ , above. You have no reason to assume that any of the twenty-six was more likely to end up there than anyone else. As far as you know, for each of them there is a  $1/26$  chance that they are the one in the last box and a corresponding  $25/26$  chance that they are among the ones in the boxes made of extra-heavy glass. This means that for each of them if you let the villain fire at the last box, there is a  $1/26$  chance that they will die and a  $25/26$  chance that they walk away completely unharmed. If you choose otherwise, on the other hand, for each hostage there is a  $1/26$  chance that they walk away unharmed (if they are the one in the last box) and a  $25/26$  chance that they walk away slightly but permanently disfigured.

I have arbitrarily assumed that this disfigurement leaves them at utility level 9.5 out of 10.<sup>19</sup> If we also assume that death leaves the hostages at “utility level” 0, then the choice situation can be represented by the following table.

|                | S <sub>1</sub> |        | S <sub>2</sub> |        | S <sub>3</sub> |        | ... | S <sub>26</sub> |        |
|----------------|----------------|--------|----------------|--------|----------------|--------|-----|-----------------|--------|
|                | A              | Others | B              | Others | C              | Others | ... | Z               | Others |
| First 25 Boxes | 10             | 9.5    | 10             | 9.5    | 10             | 9.5    | ... | 10              | 9.5    |
| Last Box       | 0              | 10     | 0              | 10     | 0              | 10     | ... | 0               | 10     |

You have two available actions (again, ignoring the option of doing nothing): “first 25 boxes” and “last box.” There are twenty-six equiprobable states of the world (S<sub>1</sub>–S<sub>26</sub>,  $p = 1/26$ ), corresponding to the twenty-six possibilities of who could be the one in the last box, where S<sub>1</sub> corresponds to the state of the world in which Ann is the one in the last box, S<sub>2</sub> to the state in which Bob is the one,

19 If you think that this is too low or too high, then you can adjust the level and change the number of hostages accordingly without affecting the basic structure of the case.

and so on. The table shows the utility levels of the hostages for each of these twenty-six states and the two available actions. For example, if you decide on “first 25 boxes” and Ann is the one in the last box, then she will be left at level 10, corresponding to full health, while the other twenty-five hostages (the “others”) will be left at level 9.5. On the other hand, if you decide on “last box” and Ann is the one in that box, she will die (“level 0”) and the others will be left unharmed at level 10. We can now calculate the expected utility for each hostage under each of the two available actions. If you choose “first 25 boxes,” the expected utility is 9.52.<sup>20</sup> If you choose “last box,” the expected utility is 9.62.<sup>21</sup> Since 9.62 is greater than 9.52, if you want to do what is in each of the hostage’s best interest, you ought to choose “last box.” If you could ask them, they would want you to do so, or if for each of the hostages there was a guardian present who is only motivated by their beloved’s interest, they would tell you to do so. Therefore, via the argument from the single-person case, *ex ante* contractualists (and proponents of *ex ante* Pareto in general) are committed to letting the villain kill the person in the last box. Note that here *ex ante* contractualism is so committed although “there is someone whom we know will die.” We know that there is a person we will willingly sacrifice—namely the person in the last box.

This fact points to an objection that could be pressed against my exposition: it is not in fact true that every hostage has a  $1/26$  chance of being the one in the last box.<sup>22</sup> At the time of decision there is a fact of the matter who the person in that box is. From this it follows that it is not actually true that choosing “first 25 boxes” is in the best interest of everyone.

I think this objection will not succeed, at least not for an *ex ante* contractualist of Frick’s kind. This is because this same objection could be pressed against someone, like Frick, who distinguishes between  $V_3$  and  $V_3^*$ , above, deeming  $V_3$  impermissible and  $V_3^*$  permissible. In both  $V_3^*$  and  $V_3$  there is a fact of the matter who the children are that are going to die. The only difference is that in  $V_3^*$  informational constraints keep us from knowing the identities of these children. The same holds for *Glass Box Villain (Unknown Victim)*. So if one thinks that we should not distinguish between *Glass Box Villain (Unknown Victim)* and *Glass Box Villain (Known Victim)* because in both cases there is a fact of the matter who is in the last box, then by parity of reasoning we also ought not to distinguish between  $V_3$  and  $V_3^*$ , because here in both cases there is also a fact of the matter who the one thousand children are for whom the vaccine will do

20  $1/26 \times 10 + 25/26 \times 9.5$ .

21  $25/26 \times 10$ .

22 This was suggested to me by Jeff McMahan and Tom Sinclair.

nothing.<sup>23</sup> As things stand, *ex ante* contractualists are committed to choose “first 25 boxes” in *Glass Box Villain (Known Victim)* and “last box” in *Glass Box Villain (Unknown Victim)*.

I think this result should worry *ex ante* contractualists, especially since, *qua* contractualists, they would be deeply committed to choose otherwise in *Glass Box Villain (Known Victim)*. The case thus lays bare the implications of the view that its proponents need to accept. These are implications many contractualists, or nonconsequentialists in general for that matter, find hard to stomach. *Glass Box Villain (Unknown Victim)* shows that *ex ante* contractualists need in some cases to be prepared to sacrifice a person’s life in order to protect many other people from a relatively minor ailment. This strikes me exactly as the kind of interpersonal aggregation that contractualism set out to avoid in the first place.

Now, *ex ante* contractualism’s proponents might be prepared to bite the bullet. They could say that the fact that the number of people affects each individual prospect (holding everything else fixed) is simply directly implied by the way *ex ante* contractualism is defined. Frick calls this “counting the numbers without aggregating.”<sup>24</sup> One could thus object to my exposition that I am implicitly assuming what I intend to show—namely that *ex ante* contractualism cannot be correct. For if one instead assumes that *ex ante* prospects are what we should be concerned with in a case like *Glass Box Villain*, then it plainly follows that we should order the villain to fire at the last box. To some extent this objection is warranted, for I am assuming that a theory that tells us to let the person in the last box be killed in *Glass Box Villain (Unknown Victim)* should strike contractualists as dubious, if not wrong. The point is that, rather than embracing this “number counting” as a welcome implication of the view, contractualists should be worried about a view that has these implications since it allows the numbers of people on each side of a binary choice to affect what we ought to do, even though the individual benefits and burdens are not affected.

The reason why most contractualists (and other nonconsequentialists) are opposed to interpersonal aggregation is because it violates what, following Rawls, has come to be called the “separateness of persons.”<sup>25</sup> According to one very strict version of this thesis the aggregation of harms across different individuals is meaningless since there is no single entity to suffer the aggregate harm. As C. S. Lewis writes:

23 This, of course, is the position I am arguing for. It is however not available to *ex ante* contractualists, as they want to distinguish between  $V_3$  and  $V_3^*$ .

24 Frick, “Contractualism and Social Risk,” 201.

25 See Rawls, *A Theory of Justice*, 167.

Suppose that I have a toothache of intensity  $x$ : and suppose that you, who are seated beside me, also begin to have a toothache of intensity  $x$ . You may, if you choose, say that the total amount of pain in the room is now  $2x$ . But you must remember that no one is suffering  $2x$ : search all time and space and you will not find that composite pain in anyone's consciousness. There is no such thing as a sum of suffering, for no one suffers it.<sup>26</sup>

This, however, is exactly what *ex ante* contractualists overlook in *Glass Box Villain (Unknown Victim)*. If we let the number of people in the boxes made of extra-heavy glass affect what we believe we ought to do, then we are overlooking the fact that the harm any of the hostages is going to suffer does not increase or decrease with that number.

Might *ex ante* contractualists respond to my argument so far by claiming that there is a principled difference in importance between saving an identified person and saving an unidentified person that I have overlooked? If so, this difference could explain why we should in fact let the villain kill the person in the last box in *Glass Box Villain (Unknown Victim)*, while we should stop him from killing Zeta in *Glass Box Villain (Known Victim)*, as well as explain why we should choose  $V_2$  over  $V_3$ , but be indifferent between  $V_2$  and  $V_3^*$ . In the following section I investigate this possibility.

### 3. THE "PRO IDENTIFIED LIVES ARGUMENT"

Many people attach greater importance to saving identified lives than to saving unidentified lives.<sup>27</sup> It is doubtful, however, that this psychological fact is of any moral relevance.<sup>28</sup> I, for one, do not think it is. It will have to be, however, in order to justify *ex ante* contractualism's way of distinguishing between  $V_3$  and  $V_3^*$ , as well as between *Glass Box Villain (Unknown Victim)* and *Glass Box Villain (Known Victim)*. Luckily for *ex ante* contractualists, Frick offers an ingenious argument to that effect. He argues that correctly applying the *ex ante* contractualist rationale to cases that are "competitive *ex ante*" yields the conclusion that we ought to prioritize identified over unidentified lives. In this section I attempt to show that this argument does not succeed.

In section 1 we saw how *ex ante* contractualism coincides with the *ex ante*

<sup>26</sup> Lewis, *The Problem of Pain*, 103–4.

<sup>27</sup> See Moore, "Caring for Identified versus Statistical Lives"; Jenni and Loewenstein, "Explaining the 'Identifiable Victim Effect.'"

<sup>28</sup> See Schelling, "The Life You Save May Be Your Own"; Brock and Wikler, "Ethical Challenges in Long-Term Funding for HIV/AIDS"; and Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability."

Pareto principle in cases in which there are actions that are in the *ex ante* interest of everyone. The principle, however, does not apply in cases that are competitive *ex ante*. Here every action that is in the interest of one group of people comes at a cost to another group of people even at the *ex ante* stage. Take the following example employed by Frick.<sup>29</sup>

*Miners:* A single miner, Jones, is trapped in a mineshaft and if we don't help him, he will die. The rescue mission, however, would be very costly. These resources could instead be used to make the mine safer for everyone working there in the future. Suppose there are 100 other people working at the mine and with the resources we would have to use on the rescue mission, we know that we could instead reduce their risk of suffering a fatal accident from 3 percent to 1 percent. What should we do?<sup>30</sup>

If we decide to let Jones die and make the mine safer for future workers, we can expect to save two workers' lives in the future instead of saving Jones's life now.<sup>31</sup> Frick argues that *ex post* contractualists here are committed to letting Jones die and saving the two other workers' lives instead.<sup>32</sup> This is because no matter what we do, the strongest individual complaints are equally strong in both cases. These are the complaints of the miners who will die when we could have prevented it, Jones in the one case and the unnamed two miners in the other case. And since under Scanlon's contractualism "numbers break ties" when the strongest complaints are equally strong on both sides, we ought to do what satisfies the greater number of strongest claims.<sup>33</sup>

Again, Frick thinks *ex post* contractualism goes wrong here. He offers his "pro identified lives argument" to show why this is so and takes this argument to provide a principled defense of the claim that we ought to prioritize identified lives over unidentified lives. The argument starts from the premise that, in general, people have a stronger claim to be saved from suffering a harm with

29 See Frick, "Contractualism and Social Risk," 212.

30 Further assume that we know that no one else but these one hundred people will ever work at the mine.

31  $100 \times 0.02$ .

32 See Frick, "Contractualism and Social Risk," 214.

33 Scanlon's "tie-breaking argument," where he draws on an argument by Frances Kamm (see Kamm, *Morality, Mortality*, 101, 114–19; Scanlon, *What We Owe to Each Other*, 229–41) is contested (see, e.g., Otsuka, "Scanlon and the Claims of the Many versus the One" and "Saving Lives, Moral Theory, and the Claims of Individuals"). However, despite the defects of this particular argument, I find it highly plausible that, when deciding between one claim on the one hand and two claims of equal magnitude on the other hand, we ought to satisfy the two claims.

certainty than to be saved from suffering that same harm with some probability  $p < 1$ .<sup>34</sup> This claim is undoubtedly correct. Suppose we have to decide between saving Ann from certain death or reducing Bob's risk of death from 3 percent to 1 percent. It is clear that we ought to help Ann in this case. Now Bob's claim to have his death risk reduced is identical, Frick continues, to each of the one hundred miners' claims in *Miners*. From this it follows that no individual miner has a stronger complaint than Jones. Coupled with the contractualist ban on interpersonal aggregation, it follows that we ought to minimize the single strongest complaint and save Jones.

Frick claims that, first, this argument provides a principled defense for the privileging of identified over unidentified lives and, second, that it also shows where *ex post* contractualism goes wrong. He claims that *ex post* contractualists are committed to the view that in *Miners* there is someone who has a stronger claim than Bob in the one versus one case. He writes, "somehow, the fact that, if we save [Jones], it is foreseeable that *someone* from the group of 100 will die in a future accident is thought to strengthen the complaint of whoever turns out to be harmed."<sup>35</sup> This, Frick argues, is an implicit appeal to interpersonal aggregation over "different possible worlds."

Regarding the first point: I think that the argument does not provide a principled defense for favoring identified lives in general, but only in a very narrow class of cases like *Miners*. It only provides a defense for favoring an identified person *when and because* that person holds a claim that is stronger than any competing claim. It, for example, does not provide a defense of the type needed to justify the *ex ante* contractualist's choices in *Glass Box Villain (Unknown Victim)*. Here, the dialectic of comparing *ex ante* claims and then satisfying the single strongest claim does not work, since here all *ex ante* claims are equally strong, as we have seen. It thus fails to provide a justification for why it is more important to save Zeta in *Glass Box Villain (Known Victim)*, than to save the unidentified person in the last box in *Glass Box Villain (Unknown Victim)*. This is because the argument does not provide a principled defense for the claim that it is more important to save an identified person rather than an unidentified person *because* that person is identified. Such a defense, however, would be needed to justify *ex ante* contractualism's verdicts in the *Glass Box Villain* cases.<sup>36</sup>

Regarding the second point, first of all, it is not clear that *ex post* contractual-

34 See Frick, "Contractualism and Social Risk," 215. See also Frick, "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives," 188–91.

35 Frick, "Contractualism and Social Risk," 217.

36 For an attempt at providing an argument to that effect, see Hare, "Should We Wish Well to

ism really is committed to letting Jones die in *Miners*. We do not *know* that two miners will die in the future if we decide to save Jones. Yes, this is the expected outcome, but, of course, it is only one of many different possible outcomes. The chance that *exactly* two miners will die is only around 27 percent.<sup>37</sup> The chance that *at least* two miners will die, so as to tip the scales in favor of letting Jones die, is around 60 percent.<sup>38</sup> No part of *ex post* contractualism commits proponents of the view to disregard these probabilities entirely. Frick assumes that they would take the expected outcome of an action and then simply act as if they knew that that expected outcome would actually eventuate. This, of course, would be a mistake. By doing so they would not be able to differentiate between cases like *Miners* and a case in which we have to decide between saving one person from certain death and saving two different people from certain death. But I do not think that anything commits them to this precarious position. Instead, they could take into account the likelihood of enough miners dying so as to outweigh Jones's claim. I take this to be the most plausible interpretation of *ex post* contractualism.<sup>39</sup> As we have seen, the likelihood of at least two miners dying is only 60 percent. So why should we just assume that *ex post* contractualists would not rescue Jones?

Second, as I have argued before, the main problem with many instances of interpersonal aggregation of harms is that any sum of weaker harms together does not constitute anything meaningful, since there is no one suffering from this aggregate harm. This, however, is not the case in *Miners*. Here, the aggregate of the many trivial harms *is* suffered by a single individual. The more people who work at the mine, the likelier it becomes that *someone* will die as a result of us not making the mine safe. This is a different kind of aggregation. Contrast this with a variation of *Miners*, where we can either save Jones or use the resources to distribute lifelong supplies of aspirin to all future miners who occasionally suffer headaches because of the stuffy air in the mine. This aggregation is more like the kind of aggregation employed by the *ex ante* contractualist in *Glass Box Villain (Unknown Victim)*. Here, as we have seen, the number of people involved has no effect on the harm that the most burdened individual has to suffer. As long as we lack an independent objection against this second, different kind of

---

All?" 267–71. I am not convinced by Hare's argument. Discussing it here, however, would lead us too far afield.

37  $100\% \times 0.02^2 \times 0.98^{98}$ .

38  $\Pr(100 \text{ deaths}) - \Pr(0 \text{ or } 1 \text{ deaths})$ .

39 I attempt to fully specify such a view elsewhere. For a similar account, see also Otsuka, "Risking Life and Limb," and Horton, "Aggregation, Complaints, and Risk," 65–66.

aggregation, I do not see why *ex post* contractualists need to be moved by this particular argument.

#### 4. PLURALISM AS A WAY OUT?

Let me now turn to the final problem with *ex ante* contractualism that I want to raise in this essay. This problem is acknowledged by Frick himself. Consider a variation of *Miners*, only now there are one thousand other miners in addition to Jones. Call this *Miners 1,000*. In this case, we would expect twenty miners to die in the future if we save Jones now. Frick submits that *ex ante* contractualism here “goes too far.”<sup>40</sup> For him it is clear that given some number of expected deaths (which could be greater or less than twenty) we ought to let Jones die. Frick concedes that this problem for *ex ante* contractualism can only be solved “by scaling back the ambitions of contractualism as a moral theory.”<sup>41</sup> He argues that in cases in which his theory is unable to yield the intuitively correct verdicts, it should be assisted by other noncontractualist principles. As a candidate, Frick suggests that we should take into account the effect an action has on people’s well-being in general. For example, in *Miners 1,000* we should take into account that there will be “a much greater loss of life” if we save Jones.<sup>42</sup>

This sounds like Frick is suggesting that the contractualist should call utilitarianism to her rescue when her theory fails her intuitions. This *ad hoc* move, however, is available to *ex post* contractualists as well. They, too, can be pluralists about interpersonal morality. Like Frick, they too can say that in some cases their theory needs to be assisted by impersonal concerns to decide what the right course of action is. I see no reason why, *prima facie*, it should seem more plausible to restrict *ex ante* contractualism in such a way than it is to restrict *ex post* contractualism in the same manner. The only difference being that it would be different cases that the theory can deal with “on its own.” In *Mass Vaccination*, for example, facing a choice between  $V_1$  and  $V_2$ , *ex post* contractualists can say the following: “In principle we ought to choose  $V_1$  here, since this minimizes the largest complaint *ex post*. However, the consequences of doing so in terms of overall well-being are too grave to be ignored. After all, if we do choose  $V_1$ , we will leave one million children with only one functioning leg for the rest of their lives. This overall loss in well-being is much greater than if one thousand children die prematurely.” If *ex ante* contractualists can legitimately resort to these

40 Frick, “Contractualism and Social Risk,” 219

41 Frick, “Contractualism and Social Risk,” 219.

42 Frick, “Contractualism and Social Risk,” 222.



impersonal reasons when their theory yields intuitively unattractive implications, the same route should be open for *ex post* contractualists as well.

However, I have misgivings about this ready resort to pluralism. As Frick himself notes, Scanlon's theory itself is already pluralist in a way.<sup>43</sup> He limits his contractualism to the domain of "what we owe to each other." We might call this domain of morality *interpersonal morality* or, following Kamm, *MI*.<sup>44</sup> However, Scanlon deems his theory to exhaust this part of morality. Frick, on the other hand, thinks that contractualism should be assisted by other principles even within this already limited domain. The question then is how valuable contractualism is as a theory above and beyond these other principles. I suspect that it is no longer very valuable. Rather, it seems unacceptably *ad hoc* and gerrymandered to fit a very narrow class of cases. Whenever we look beyond this narrow class of cases and the theory fails to yield the right result, its proponent can resort to pluralism. Leaving the theory open in this way, however, limits its value. It means that the theory has too many free parameters, limiting its predictive power and testability, thereby putting into doubt its value as a standalone moral theory.

Moreover, even granting this pluralistic approach, it will not get *ex ante* contractualism around the implications of the *Glass Box Villain* case in section 2. Utilitarianism here pulls in the same direction as *ex ante* contractualism. All other things equal, the overall aggregate value of an outcome where twenty-five people are spared a 0.5-unit decrease in utility is higher than the value of an outcome where one different person is spared a ten-unit decrease in utility. So even if the misgivings I have with the pluralistic approach are unwarranted, this problem remains.

## 5. CONCLUSION

In this article I have challenged Johann Frick's *ex ante* contractualism. I argued that adopting the view leads to implications contractualists will find hard to stomach. This has become especially vivid in *Glass Box Villain (Unknown Victim)*, where *ex ante* contractualists are committed to sacrificing one person in order to save twenty-five different people from relatively minor harm. I have argued that this is an instance of the kind of interpersonal aggregation of harms that contractualists sought to avoid in the first place. I also argued that this kind of aggregation is more troublesome than the kind of aggregation Frick accuses *ex post* contractualists of. In connection to this last point I have argued that

43 See Frick, "Contractualism and Social Risk," 220n47.

44 Kamm, *Intricate Ethics*, 455–90.

Frick's argument for the principled priority of identified over unidentified lives also fails, because it can only account for *ex ante* contractualism's verdict in a very narrow class of cases. Finally, I have argued that Frick's resort to pluralism is *ad hoc* and further unable to block some of the unwelcome implications of the view. I conclude that, if there is no other way of developing *ex ante* contractualism that does not run into these problems, contractualists ought to be concerned with the probability that harm could befall someone, rather than with the probability that harm could befall a specific person. For contractualists, a suitably amended *ex post* approach is better equipped to honor this commitment.<sup>45</sup>

Balliol College, University of Oxford  
 korbinian.rueger@balliol.ox.ac.uk

#### REFERENCES

- Ashford, Elizabeth. "The Demandingness of Scanlon's Contractualism." *Ethics* 113, no. 2 (January 2003): 273–302.
- Brock, Dan W., and Daniel Wikler. "Ethical Challenges in Long-Term Funding for HIV/AIDS." *Health Affairs* 28, no. 6 (November/December 2009): 1666–76.
- Dougherty, Tom. "Aggregation, Beneficence, and Chance." *Journal of Ethics and Social Philosophy* 7, no. 2 (May 2013): 1–19.
- Fleurbaey, Marc, and Alex Voorhoeve. "Decide as You Would with Full Information! An Argument Against *ex ante* Pareto." In *Inequalities in Health: Concepts, Measures, and Ethics*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 113–29. Oxford: Oxford University Press, 2013.
- Frick, Johann. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43, no. 3 (Summer 2015): 175–223.
- . "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives." In *Identified versus Statistical Lives: An Interdisciplinary Perspective*, edited by I. Glenn Cohen, Norman Daniels, and Nir Eyal, 182–203. Oxford: Oxford University Press, 2015.
- Fried, Barbara H. "Can Contractualism Save Us from Aggregation?" *Journal of Ethics* 16, no. 1 (March 2012): 39–66.

45 I thank Ralf Bader, Hilary Greaves, Joe Horton, Kacper Kowalczyk, Benjamin Lange, Jeff McMahan, Johanna Privitera, Tom Sinclair, Bastian Steuwer, and an anonymous reviewer of this journal for helpful comments on earlier versions of this article.

- Hare, Caspar. "Should We Wish Well to All?" *Philosophical Review* 125, no. 4 (October 2016): 451–72.
- Horton, Joe. "Aggregation, Complaints, and Risk." *Philosophy and Public Affairs* 45, no. 1 (Winter 2017): 54–81.
- Jenni, Karen E., and George Loewenstein. "Explaining the 'Identifiable Victim Effect.'" *Journal of Risk and Uncertainty* 14, no. 3 (May/June 1997): 235–57.
- Kamm, F. M. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: Oxford University Press, 2007.
- . *Morality, Mortality*, vol. 1, *Death and Whom to Save from It*. Oxford: Oxford University Press, 1993.
- Kumar, Rahul. "Risking and Wronging." *Philosophy and Public Affairs* 43, no. 1 (Winter 2015): 27–51.
- Lewis, C. S. *The Problem of Pain*. New York: Collins/Fontana Books, 1957.
- Lewis, David. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, edited by Richard C. Jeffrey, 83–132. Lanham, MD: University Press of America, 1980.
- Moore, Randall F. "Caring for Identified versus Statistical Lives: An Evolutionary View of Medical Distributive Justice." *Ethology and Sociobiology* 17, no. 6 (November 1996): 379–401.
- Otsuka, Michael. "Risking Life and Limb: How to Discount Harms by Their Improbability." In *Identified versus Statistical Life: An Interdisciplinary Perspective*, edited by I. Glenn Cohen, Norman Daniels, and Nir Eyal, 77–93. Oxford: Oxford University Press, 2015.
- . "Saving Lives, Moral Theory, and the Claims of Individuals." *Philosophy and Public Affairs* 34, no. 2 (March 2006): 109–35.
- . "Scanlon and the Claims of the Many versus the One." *Analysis* 60, no. 3 (July 2000): 288–93.
- Rawls, John. *A Theory of Justice*. Oxford: Oxford University Press, 1999.
- Reibetanz-Moreau, Sophia. "Contractualism and Aggregation." *Ethics* 108, no. 2 (1998): 296–311.
- Scanlon, T. M. "Reply to Zofia Stemplowska." *Journal of Moral Philosophy* 10, no. 4 (2013): 508–14.
- . *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.
- Schelling, Thomas C. "The Life You Save May Be Your Own." In *Problems in Public Expenditure Analysis*, edited by Samuel B. Chase, 127–62. Washington, DC: Brookings Institution, 1968.
- Voorhoeve, Alex. "Why One Should Count Only Claims with Which One Can Sympathize." *Public Health Ethics* 10, no. 2 (July 2017): 148–56.