

# PREFERENCE COALITIONS AND THE ETHICS OF AGGREGATION

Russell McIntosh

## ABSTRACT

Some forms of aggregation, like intrapersonal aggregation, seem easier to justify than others. What conditions, in general, make aggregation easier to justify? Although many factors are plausibly relevant, this paper defends the Respectable Preference-Based Principle, on which aggregation is easier to justify when the affected parties respectably prefer it. Preferences are respectable, I argue, when they are informed, sufficiently reasonable, and sufficiently authentic. The view that emerges justifies aggregation by appealing not to impersonal value, but to personal value *for* the affected parties. This has some surprising upshots: intrapersonal aggregation is only sometimes fully justifiable; interpersonal aggregation can be even *easier* to justify than intrapersonal aggregation; and, in contrast with the popular idea that deontology covers people and utilitarianism covers animals, interanimal aggregation is in one important way *harder* to justify than interpersonal aggregation. The resulting view suggests a new way to draw moral lines, based not on the separateness of persons, but on the separateness of what I call “preference coalitions.”

Keywords: aggregation, animal ethics, directed justification, preferences, separateness of persons

## 1. THE METAPHYSICAL PRINCIPLE

Some actions that promote aggregate welfare—summed burdens and benefits—seem easier to justify than others. For example, withholding benefits from a person now to promote her aggregate welfare over time seems easier to justify than withholding benefits from a person to promote the aggregate welfare of many others. *Intrapersonal* aggregation seems easier to justify than *interpersonal* aggregation.

What conditions, in general, make aggregation easier to justify? The dominant view is

*Metaphysical Principle.* Aggregation of benefits to A and burdens to B is easier to justify when A and B are the *same* person or animal.<sup>1</sup>

---

<sup>1</sup> Philosophers that would accept the Metaphysical Principle include Anscombe (1967), Doggett (2013), Foot (1985), Fischer (2021, 2024), Frick (2015), Gauthier (1963: 123–127), Kamm (1993, 2007), Korsgaard (2019), Kumar (2001), Lazar (2018), Munoz-Dardé (2005), Nagel (1970: 133–142), Nozick (1974), Otsuka (2006), Rawls (1971/1999), Scanlon (1996), Setiya (2014), (arguably) Sidgwick (1874), Steuwer (2020), Taurek (1977), Temkin (2011), Thomson (2008),

Where an action is *easier to justify* if it is, if wrong, less seriously wrong or, if permissible, more praiseworthy and choiceworthy.<sup>2</sup> For illustration, consider

**Child's Procedure.**

We can either

- (1) put a child through a painful medical procedure so she will receive many small benefits later in life, or
- (2) put a child through a painful medical procedure so many strangers will receive small benefits.

By the Metaphysical Principle, 1 is easier to justify than 2. This seems like the right result: intrapersonal aggregation seems easier to justify than interpersonal aggregation. Indeed, the Metaphysical Principle just restates that intuition.

Many accept the Metaphysical Principle. It is a natural interpretation of Rawls's criticism of interpersonally aggregative theories that they fail to respect the separateness of persons.<sup>3</sup> It is also suggested by Nozick's claim that

Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm [...]. But there is no *social entity* with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits others. Nothing more.<sup>4</sup>

And it is in the spirit of Taurek's claim that

The discomfort of each of a large number of individuals experiencing a minor headache does not add up to anyone's experiencing a migraine.  
In such a trade-off situation as this we are to compare your pain or your

---

Voorhoeve (2014), Wallace (2019), and Zhang (2024). For philosophers skeptical that the metaphysical separateness of persons makes interpersonal aggregation harder to justify, see Brink (1993), Halstead (2016), Hirose (2013), Horton (2018), Parfit (1984, part 3), and Yetter Chappell (2015).

<sup>2</sup> See Hurka 2019 for discussion of the idea that actions can be more or less seriously wrong. For example, breaking a promise to meet for lunch is less seriously wrong than murder.

<sup>3</sup> Rawls 1971/1999: 23–24.

<sup>4</sup> Nozick 1974: 32–33, emphasis original.

loss, not to our collective or total pain, whatever exactly that is supposed to be, but to what will be suffered or lost by *any given single one of us*.<sup>5</sup>

Each of Taurek's headache patients values the relief of their pain. The relief of each headache is personally good *for* each patient. But, Taurek's thought is, the *aggregate* of these headaches is only *impersonally bad*: bad from no particular person's point of view. The Metaphysical Principle is attractive because it's sensitive to this difference between value for particular people and impersonal value—value not valued *by* anyone in particular, good not good *for* anyone in particular. The justificatory force of value-for doesn't reduce to that of impersonal value. In Child's Procedure, 1 & 2 create the same amount of impersonal value. But 1 is easier to justify: we can better justify our actions *to* Child by trading off her bads with *her* goods, rather than by trading off her bads with *others'* goods.

Despite its attractions, the Metaphysical Principle cannot be the complete story of justifiable aggregation. It fails to explain why some forms of intrapersonal aggregation are easier to justify than others. Consider

**Dave & Ursula.**

While Divided Dave is young, he cares especially about his young self and considers his old self a stranger. While Dave is old, he cares especially about his old self, and considers his young self a stranger. Throughout Unified Ursula's life, she cares equally about all its parts. We can either

- (1) burden young Dave and significantly benefit old Dave, or
- (2) burden young Ursula and significantly benefit old Ursula.

Intuitively, 2 is easier to justify than 1. If we do 1, young Dave could object that we are not doing what he prefers. But if we do 2, young Ursula and old Ursula would unanimously approve. However, the Metaphysical Principle cannot detect this difference in justifiability. The Metaphysical Principle only makes comparisons between intra- and interpersonal aggregation—it cannot distinguish between different cases of intrapersonal aggregation.

But even in comparisons between intra- and interpersonal aggregation, the Metaphysical Principle gives misleading results. Consider

---

<sup>5</sup> Taurek 1977: 308, emphasis original.

**Parent & Child.**

We can either

- (1) burden Parent now and give Parent a larger benefit later, or
- (2) burden Parent now and give Child a larger benefit later,

where Parent and Child would each prefer 2.

Intuitively, 1 is harder to justify—for starters, it seems objectionably paternalistic to override Parent’s preference to sacrifice. Even if the Metaphysical Principle is ultimately right that 1 is *in one way* easier to justify, the principle is insensitive to some ethically overriding feature that we should try to identify more precisely.

Can we explain what the Metaphysical Principle is missing? Can we explain why some forms of intrapersonal aggregation are easier to justify than others, as in Dave & Ursula, and why interpersonal aggregation can be easier to justify than intrapersonal aggregation, as in Parent & Child?

The Metaphysical Principle ignores that it matters not only *who I am* but *what I care about*. It construes value-for *narrowly*, in terms of welfare, rather than *widely*, in terms of what we care about. The fact that Ursula cares about her future self, while Dave does not, affects how we can justifiably treat her. The fact that Ursula’s aggregate welfare is valued *by someone* lends it a special significance. Ursula’s stable concern for her own aggregate welfare creates univocality: no part of her would dissent were we to choose to benefit her in the aggregate, even if doing so burdens her young self. Similarly, the fact that Parent cares about Child creates a univocality that a focus on the metaphysical separateness of persons obscures. In contrast, although each part of Dave’s life is cared about by someone, there’s no single perspective from which *all* of Dave’s life matters equally. Dave’s fractured concerns create an ethical conflict: we can either do what young Dave would want or what old Dave would want, but not what both want.

In general, actions that promote aggregate welfare are easier to justify when the affected parties prefer those actions. That’s a first pass of the view I’ll defend here. I’ll argue for it in section 2 and refine it in section 3. My view explains why some forms of intrapersonal aggregation (like burdening and benefitting Ursula) are easier to justify than others (like burdening and benefitting Dave). It also explains why some cases of interpersonal aggregation (like burdening Parent and benefitting Child) can be easier to justify than some cases of intrapersonal aggregation (like burdening Parent and benefitting Parent). After discussing implications in section 4 and further questions in section 5, I’ll return to the question of what conditions, in general, make aggregation easier to justify in section 6. My view suggests a new way to think about aggregation generally, as organized not around the separateness of persons, but around the separateness of what I will call “preference coalitions.”

## 2. THE PREFERENCE-BASED PRINCIPLE

Preferences can justify aggregation. To my knowledge, no one has explicitly defended, or even entertained, this simple idea.<sup>6</sup> Philosophers writing about constraints on aggregation have focused on the separateness of persons. But they have packed into the notion of separate persons the separateness of their concerns. Rawls claims, for instance, that utilitarianism's impartial spectator device fails to respect the separateness of persons by "identif[ying] with and *experienc[ing] the desires of others* as if these desires were his own."<sup>7</sup> Voorhoeve claims that "the non-aggregative approach involves imaginatively placing oneself, one person at a time, in the position of each person who has a claim and *viewing the situation through her eyes*."<sup>8</sup> And Steuwer claims that "respecting the separateness of persons requires *taking into consideration each person's point of view* separately."<sup>9</sup> The metaphors of experiencing another's desires, viewing through her eyes, and considering her point of view evoke not the significance of separate spatial locations, but rather the significance of separate evaluative focal points, separate loci of concerns, separate profiles of preferences. We can think of these as *points of view*. Rawls's, Voorhoeve's, and Steuwer's claims, which represent a prominent strand of moral philosophy, draw on the idea that the separateness of persons matters because different people have different points of view from which their concerns have a special significance.

But we should not conflate the separateness of persons and the separateness of points of view. Persons and points of view come apart. Two people can share a shallow point of view merely by sharing a preference, and one person's point of view fragments when their preferences change—or change in certain ways, e.g., without a change in information. As it happens, distinct people and distinct points of view often walk together. But it's worth seeing how far points of view alone can take us, without using the intermediate concept of the person.

As a first-pass, let's consider

*Preference-Based Principle.* Aggregation of benefits to A and burdens to B is easier to justify when A and B prefer that we promote A and B's aggregate welfare.<sup>10</sup>

---

<sup>6</sup> However, a common methodology in the theory of personal identity may implicitly assume this idea. I discuss this in 6.1.

<sup>7</sup> Rawls 1971/1999: 27, emphasis mine.

<sup>8</sup> Voorhoeve 2014: 5, emphasis mine.

<sup>9</sup> Steuwer 2020: 3, emphasis mine.

<sup>10</sup> My principle obviously resembles a *weak Pareto* principle, which says we should do what's better for everyone, if we can. However, people can prefer what's not better for them, as when Parent prefers that Child benefits. And people can be benefited by outcomes they don't prefer—otherwise the ethical question of paternalistic intervention is moot. So the Preference-Based Principle is meaningfully different from weak Pareto. However, if the reader thinks that what's better for me is just what I prefer, then she can read this paper in a different light, as an argument that treating people in ways they

A *preference*, as I use the term, is a disposition to choose some action or outcome. We can say that A and B form a *preference coalition* when they prefer the same outcome. Several people can form a preference coalition, as when a group of friends prefer that they eat at the Mexican restaurant. And a single person can form two or more: Parfit's Russian nobleman, the young socialist who becomes an old conservative, forms at least two preference coalitions. The young socialist would prefer that the estate go to the peasants; the old conservative would prefer to keep the estate for himself. (Parfit 1973: 145–146) So preference coalitions come apart from people and the metaphysics of personal identity.

The Preference-Based Principle correctly implies that promoting Ursula's aggregate welfare is easier to justify than promoting Dave's: Ursula, but not Dave, forms a preference coalition that univocally prefers that we promote her aggregate welfare, even at the expense of young Ursula. Similarly, the Preference-Based Principle correctly implies that benefitting Child is easier to justify than benefitting Parent: Parent and Child form a preference coalition that univocally prefers that we benefit Child. The Preference-Based Principle, unlike the Metaphysical Principle, is sensitive to what actually matters to people.

The Preference-Based Principle says that an action that promotes aggregate welfare is easier to justify when the affected parties prefer that action. But, to clarify, the principle does not require that the affected parties prefer the action *on the grounds* that the action promotes aggregate welfare. The Preference-Based Principle focuses on the *content* of preferences, not the grounds. Parent may prefer that we benefit Child because Parent cares about Child's aggregate welfare as such. But suppose Parent is skeptical about the very idea of welfare and instead prefers that we benefit Child on alternative grounds: say, Parent believes she's obliged by her role to have such a preference. The Preference-Based Principle still counts such a preference as a preference for an action that promotes aggregate welfare, even if it's not a preference for welfare-promotion as such.

Its verdicts in Dave & Ursula and Parent & Child make the Preference-Based Principle attractive. But we can argue more generally for the Preference-Based Principle as follows.

1. An action  $\phi$  is easier to justify *simpliciter* if it's easier to justify to each affected party.
2. An action  $\phi$  is easier to justify to a party if  $\phi$ -ing satisfies her preferences.
3. Therefore, an action  $\phi$  is easier to justify if  $\phi$ -ing satisfies the preferences of each affected party. (from 1 and 2)
4. Therefore, *Preference-Based Principle*: Aggregation of benefits to A and burdens to B is easier to justify when A and B prefer that we promote A and B's aggregate welfare. (from 3)

---

prefer (and thus benefitting them) can require some underappreciated and surprising implications, especially those I discuss in section 4.

The action happens in 1 and 2, so I'll explain why we should accept them.

Why accept 1? Contractualists will be especially amenable to 1. Contractualists say that an action is permissible if and only if it is justifiable to each person.<sup>11</sup> They thus emphasize the importance of *directed justification*: justification directed toward specific people. In the same spirit, they might say that an act is easier to justify if it's easier to justify *to* each affected party. But even non-contractualists should accept this rather weak claim: 1 says only that easier justifiability-to-each makes for easier justifiability *simpliciter*, not that justifiability-to-each is a *necessary* condition on justifiability *simpliciter*.<sup>12</sup>

Why accept 2? If you prefer that I  $\phi$ , and instead I  $\psi$ , you might thereby have an objection to my  $\psi$ -ing.<sup>13</sup> That objection need not be decisive—it might be outweighed or defeated by powerful objections to my  $\phi$ -ing—but it is an objection nonetheless. However, when I  $\phi$ , you no longer have any such objection. Instead, I can offer a new, compelling justification of my action to you. I can say to you, “Given the things that you, yourself, care about, you should accept  $\phi$ -ing as the best action available to me. *By your own lights*  $\phi$ -ing is the thing to do.” Such by-your-own-lights justifications are, on the face of it, especially powerful. Indeed, it's hard to see how you could sincerely complain when I offer you such a justification.<sup>14</sup>

Two considerations speak in favor of the power of by-your-own-lights justifications.

First, it seems we have general reasons to satisfy others' preferences. Philosophers are unlikely to agree on any particular source of such reasons, but there are many plausible candidates: we could respect others' autonomy by helping them to live lives in accordance with their own values<sup>15</sup>; deferring to others' preferences could symbolically express respect toward them, and overriding their preferences on a matter primarily affecting them could be belittling<sup>16</sup>; others' desires can give us reasons when they have authority, even, minimally, the authority of another free and rational agent<sup>17</sup>; and doing what

---

<sup>11</sup> See Scanlon 1998, Wallace 2019. On the classic contractualist view, an act is justifiable to a person if they could not reasonably reject a principle permitting it. But see Bourguignon 2024 for an act-based alternative.

<sup>12</sup> Of course, some accept a view of directed justification on which there is no substance to directed justification beyond justification *simpliciter*. Some might say, for example, that I can justify failing to save one person from death by appealing to the impersonal goodness my action would achieve. Such people would accept a trivial reading of 1, but reject 2. Nagel (1972: 137) entertains the possibility that such “You understand, I have to leave you to save the other” justifications can succeed as directed justifications. Parfit (2011: 371–403) avows the success of appeals to impersonal value as directed justifications. But such views fit uneasily with the starting idea that some forms of aggregation seem easier to justify than others.

<sup>13</sup> This claim is most plausible for certain kinds of preferences, as I discuss in the next section.

<sup>14</sup> Again, this claim will be most plausible for certain kinds of preferences.

<sup>15</sup> As would follow from Enoch (2017)'s view of autonomy as non-alienation.

<sup>16</sup> See Scanlon (1998: 253).

<sup>17</sup> See Darwall (2001).

another would want can both respond to and strengthen our valuable relationships.<sup>18</sup> The force of by-your-own-lights justifications comes, in part, from the force of these underlying reasons to satisfy others' preferences.<sup>19</sup>

Second, on Frick (2016)'s plausible view, directed justification is a kind of *joint deliberation*: for B to successfully justify to A that A ought to  $\phi$ , this must be (part of) the correct answer to the question "what do *we*, A and B, have reason to do?" Frick puts this account of directed justification to work to explain the speaker-relativity of justification.<sup>20</sup> But I want to set aside the speaker-relativity of justification. Rather, I want to explain the power of by-your-own-lights justifications. Respectful joint deliberation, I submit, requires taking seriously our deliberative partner's preferences, which we can do by offering by-your-own-lights justifications. Consider

**Car vs Train.**

Paulina, Reza, and Denzel are deciding how to get to the party. Paulina would prefer to travel in a more environmentally friendly way. Reza says, "The train emits less pollution, and Paulina cares a great deal about that, so we should take the train." Denzel says, "Psh, are we really going to give the environment priority over people? But let's take the train anyway—I don't want to waste time parking."

By taking her preference seriously in their joint deliberation, Respectful Reza treats Paulina more respectfully than does Disrespectful Denzel.<sup>21</sup> As a species of joint deliberation, directed justification inherits this constraint on how we can treat our partner's preferences: when we justify ourselves to another, we should take seriously how their preferences are satisfied (or not) by our actions. All else equal, it will be easier to justify our actions to others when our actions satisfy their preferences.

<sup>18</sup> See McIntosh (ms.).

<sup>19</sup> We might also have reasons to promote others' welfare, which may be constituted in part by desire-satisfaction (as argued in Heathwood 2017). However, the weightier our welfare-based interests in desire-satisfaction, the more the Metaphysical Principle drifts toward the Preference-Based Principle. Thus, if there are weighty such interests, that only bolsters my first-order ethical conclusions. However, because it is helpful presentationally to keep the Metaphysical and Preference-Based Principles distinct, I set aside such reasons.

<sup>20</sup> For example, Frick's account purports to explain why your friend can justifiably advise you by saying "You should give Jones \$20, because otherwise he will kill himself," but Jones himself *cannot* justifiably advise you by saying "You should give me \$20, because otherwise I will kill myself." If you and Jones are jointly deliberating, you can respond to his proposed action of give-Jones-\$20-and-he-doesn't-kill-himself with the compelling counteroffer *don't-give-Jones-\$20-and-he-still-doesn't-kill-himself*. In a context of joint deliberation, Jones cannot deny that it is up to us, together, to decide whether he kills himself in response to his nonreceipt of \$20.

<sup>21</sup> Taking another's preference seriously requires a modally robust sensitivity to their preference. Reza's preference does not merely *align* with Paulina's. Rather, his preference is itself sensitive to Paulina's preference—were her preference to change, his would too.

However, you might object to 2 on these grounds: actions are justifiable to a person not by satisfying their preferences, but by benefitting them. You might think that the best way to justify a burden to a person is by *compensating* that very person—not *restitutively*, to make up for wrongdoing, but *distributively*, to make the burdened person whole, ideally as well off she would have been without the burden. Given that we must burden Parent, we can most easily justify our action to her by compensating her. We cannot compensate Parent by benefitting Child, the thought goes.<sup>22</sup>

This objection assumes that compensation must be made *narrowly* in terms of the burdened party's welfare. But note that a parent's sacrifice can be a genuine sacrifice while also seeming *worth it* to her, given what she cares about: we might say that, while she is not narrowly compensated in terms of her welfare, she is *holistically* compensated by benefits to other things that matter to her. By benefitting Child, we do not compensate Parent in the narrow sense for the burden that she suffers, but we do holistically compensate her. This is a valuable form of compensation: holistic compensation, unlike narrow compensation, repays someone in the coin that her own evaluative faculties minted.

So we should accept 1 and 2: an action  $\phi$  is easier to justify if it is easier to justify to each affected party, and it is easier to justify  $\phi$ -ing to a party when  $\phi$ -ing satisfies her preferences. It follows directly from these claims that  $\phi$ -ing is easier to justify when  $\phi$ -ing satisfies the preferences of everyone affected. Then, when we plug in a preference for the aggregation of burdens and benefits to the affected parties, we get the Preference-Based Principle: aggregation of benefits to A and burdens to B is easier to justify when A and B prefer it.

### 3. RESTRICTING THE PREFERENCE-BASED PRINCIPLE

The argument for premise 2 might give us pause. Not all preferences are reason-giving or suited to play a role in directed justification. Not all by-your-own-lights justifications are equally powerful. A preference is just a disposition to choose. But dispositions to choose can be ill-informed, strange, or wicked: we might be disposed to choose a certain way because we've been misled, because we're afflicted by a brute urge, or because we're evil. Justifications that appeal to such preferences seem weak. Let's call the preferences apt to be reason-giving and suited to play a role in directed justification *respectable preferences* (setting aside any political implications of the term): preferences that, even if you don't understand them, you have reasons to respect. We can then favor a restricted argument from directed justification and a correspondingly restricted Preference-Based Principle:

*Respectable-Preference-Based Principle.* Aggregation of benefits to A and burdens to B is easier to justify when A and B respectably prefer that we promote A and B's aggregate welfare.

---

<sup>22</sup> This compensation-based rationale for defending the ethical significance of the separateness of persons is suggested by Parfit (1984: 337).

Although I don't want to dismiss the Preference-Based Principle entirely—some readers may find the idea that some preferences are more respectable than others unappealing—I think the Respectable-Preference-Based Principle is the most plausible version of the principle, and it's the version I stand by.

I've identified respectable preferences by their moral theoretical function—we have reasons to satisfy them and they constrain directed justification. But that's not yet to commit to any substantial, first-order claims about which preferences are respectable. To get an idea on the table, I submit that at least informed, sufficiently reasonable, and sufficiently authentic preferences are respectable. I'll explain these in turn.

First, consider preferences *informed* by the relevant descriptive facts. Suppose you prefer to have this drink. But, if you knew the drink was poisoned, you'd prefer *not* to drink it. I'd have reason to respect the informed preference, but not your uninformed one.

Second, consider *sufficiently reasonable* preferences. A preference is *reasonable*, as I use the term, if one has sufficient reasons for it. Reasonableness thus departs from rationality, where your preference is *rational* if it coheres with your other mental states.<sup>23</sup> For example, a cruel father's preference that his child break all her limbs may be rational—it need not conflict with any of the father's other mental states. But the preference is unreasonable: the fact that it would hurt is a powerful reason not to prefer that the child break all her limbs.

I've also identified the concept of a reasonable preference merely by its function. I'll mention four substantial views of which preferences are reasonable, to give a sense of the landscape. On a *Humean View*, all preferences are trivially reasonable, because there are no reasons for preferences at all.<sup>24</sup> I've already registered disagreement with this view, but it's an option. The Humean View effectively removes the reasonableness filter on respectable preferences. On an *Egoist View*, the only reasonable preferences are self-interested ones.<sup>25</sup> If the Egoist View is right, it would be reasonable to prefer, and unreasonable not to prefer, that I get the last slice of cake (supposing that would be good for me), even if my friend came to the party late and hasn't yet had a slice. The Egoist View and the Respectable-Preference-Based Principle would together constitute an aggregation-skeptical view, much like the Metaphysical Principle, on which we can never fully justify benefiting one person at the expense of another to the one who loses out. We could instead accept a *Utilitarian View*, on which the

---

<sup>23</sup> Compare Kolodny 2005: 509–10 on objective and subjective rationality. My *reasonableness* corresponds to Kolodny's *objective rationality*. My *rationality* corresponds to Kolodny's *subjective rationality*.

<sup>24</sup> See Hume's claim that "Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger." (Hume 1738/2000: 2.3.3.6) For a competing view, see Parfit (2011: 43–57). For a consonant view, see Street (2009).

<sup>25</sup> Egoism and utilitarianism are generally characterized as ethical theories meant to give criteria of permissible action, as in, e.g. Sidgwick 1874/1981. But note that the views I give here are not theories of permissible action, but theories of reasonable preference.

only reasonable preferences are preferences for the maximization of aggregate welfare. If the Utilitarian View is right, it's reasonable to prefer, and unreasonable not to prefer, that my savings go to an effective charity. The Utilitarian View and the Respectable-Preference-Based Principle would together constitute an aggregation-friendly view, on which aggregation-friendly preferences always speak in favor of maximizing aggregate welfare.

These extreme views are, I think, implausible. The most plausible view is instead the *Moderate View*, on which some preferences are objectively unreasonable, but individuals have sufficient reason for several preferences, including preferences concerned with things besides themselves and aggregate welfare. For example, it's plausibly unreasonable not to prefer that our selves, friends, families, relationships, projects, and cultures survive into the future, all else equal. It's unreasonable to be indifferent between a healthy parent's survival and her death: we have strong reasons to prefer her survival. And it's plausible that agents have discretion to prefer what they wish, within certain bounds. I have no knock-down argument for the Moderate View, but I will assume it going forward.

The reasonableness of a preference plausibly depends not only on its content but on its grounds. For example, it's reasonable to prefer my partner get the promotion on the grounds that she deserves recognition for her hard work. On the other hand, it's unreasonable to prefer that she get the promotion on the grounds that the new position would disrupt her sleep, or on the grounds that the position title has an even number of letters.

I submit that not only reasonable preferences but *sufficiently* reasonable preferences can be respectable. As long as others' preferences are not too unreasonable, we should act as if they are reasonable. In everyday life, there should be a strong presumption that others' preferences are reasonable.<sup>26</sup> This is so for three reasons. First, in this case, false negatives are more hazardous than false positives: the objection to disrespecting a reasonable preference is more severe than the objection to respecting a moderately unreasonable preference, so we should be more careful to avoid the former error. We should then give others a wide berth, even at the risk of respecting some unreasonable preferences. Second, investigating whether others' preferences are reasonable is intrusive. We should trust others to sort out their own preferences, rather than interrogating them about the grounds for their preferences. Third, it is important for the preference-bearer's sense of security that she know she has a range within which her preferences should be respected by others, even if they turn out to be unreasonable. This gives us a reason to respect some preferences even when we know that they are unreasonable.<sup>27</sup>

---

<sup>26</sup> This presumption may be weaker for the other animals. I discuss animals in 4.3.

<sup>27</sup> Preference bearers can lose this presumption of reasonableness. For example, if I have a moderately ill-informed preference, but I've egregiously failed to do my duty to uncover the mistake (I've covered my ears, or hidden the newspapers, say), then you could reasonably say to me, "Look, I'd be happy, in general, to respect your preferences. But, in this case, you're being willfully ignorant. You really ought to know better. So I won't be respecting your preference. I hope you understand."

Third, consider *sufficiently authentic* preferences. A preference is *authentic*, intuitively, if the preference-bearer identifies with it, endorses it, or is not alienated from it. We can talk of authentic preferences as grounded in deep commitments. For example, a deep commitment to my mother might ground an authentic preference that she have a nice morning. In contrast, an unwilling addict, to borrow Frankfurt's example, might have an inauthentic preference to take a drug, in the sense that she doesn't endorse that preference or see it as expressive of who she really is. We don't have to adopt a substantive view of what makes preferences authentic to understand the contours of the Respectable-Preference-Based Principle. But, again, to get a sense of the landscape, we can keep in mind Frankfurt (1971)'s view, on which authentic preferences are supported by higher-order preferences, and Watson (1975)'s view, on which authentic preferences are supported by normative judgments. I add "sufficiently" for similar reasons to those I mentioned above.<sup>28</sup>

In sum, we should accept a restricted version of the Preference-Based-Principle that applies only to respectable preferences, and we should say that a preference is respectable at least if it is informed, sufficiently reasonable, and sufficiently authentic. To be clear, I'm offering sufficient conditions on respectable preferences, not necessary ones: it may be that, ultimately, some uninformed preferences are still respectable, for example. But I will proceed cautiously, making do only with this narrow conception of respectable preferences.

Before we move on to implications, let me make one more clarification. The Respectable Preference-Based Principle says that we should honor certain preferences when individuals have them. But it does not say that we should act as if those preferences exist when individuals do *not* have them. We can imagine that, as we are deliberating, affected parties hand us a profile of their preferences. The Respectable Preference-Based Principle tells us we can strike through and ignore non-respectable preferences. But it does not say that we should *write in* preferences that do not exist. An extended principle might include this claim, but that principle would require new argument. The above argument appeals to the idea that an action is easier to justify to someone if it satisfies her actual (respectable) preferences. Much less obvious is the idea that an action is easier to justify if it satisfies a (respectable) preference that an affected party *could* have, whether or not she actually has it.<sup>29</sup>

---

<sup>28</sup> Why not say that respectable preferences can be *sufficiently* informed? My thought is that our different values call for respect in a way that our different descriptive beliefs generally do not: others have reasons to respect my misguided environmental values, say, but not my misguided beliefs about the effectiveness of some environmental intervention. Of course, our descriptive beliefs are often influenced by our values, so these will be difficult to disentangle.

<sup>29</sup> The Respectable-Preference-Based Principle implicitly applies only to all-things-considered preferences. However, the argument in this section and the previous suggests a more general principle: that aggregation is easier to justify when the affected parties *to any degree* respectably prefer it, where the difference in ease of justification is a function of the preference's strength. For example, suppose A in one way prefers that I  $\phi$ , but all-things-considered prefers that I  $\psi$ . This more general principle would imply that A's preference for  $\phi$ -ing makes  $\phi$ -ing in one way easier to justify, even if A's stronger preference for  $\psi$ -ing makes a larger difference to the ease of justifying  $\psi$ -ing.

## 4. IMPLICATIONS OF THE RESPECTABLE PREFERENCE-BASED PRINCIPLE

### 4.1 *Intrapersonal aggregation is not always justified*

As far as the Respectable-Preference-Based Principle is concerned, intrapersonal aggregation is not always justified. This is clear from the Dave & Ursula case: intrapersonal aggregation is easier to justify for Unified Ursula than for Divided Dave, since only Unified Ursula forms a preference coalition that univocally prefers that we benefit her as a whole.

You might think that, since Divided Dave is an extreme, unrealistic case, this doesn't tell us much about how to treat actual people. We are more like Ursula than like Dave. However, even if we are more like Ursula, we depart from Ursula's unity in ways that are respectable, or, at least, in ways whose respectability is a matter for substantial debate. I'll mention three such departures from unity.

First, many of us are *future-biased*: we care more about the future than the past. Consider an extreme case of future bias, in which Jones consistently throughout his life cares only about his future self. If we had to choose between giving Jones a large benefit at age twenty and a small benefit at age eighty, we would face *coalitional conflict*: different preference coalitions would prefer different actions. Jones-up-to-twenty would prefer the large benefit at age twenty. Jones-twenty-to-eighty would prefer the small benefit later. And Jones-eighty-plus would be indifferent. I defer discussion of options for addressing coalitional conflict in general to the next section. But this clarifies the framework the Respectable Preference-Based Principle gives us for reasoning about cases of future-bias, even though typical cases are less extreme than Jones's.

Second, many of us have *proximity-bias*: we care more about events temporally close to us. Consider an extreme case of proximity bias, in which Smith lives only day-to-day, caring nothing about his yesterdays or tomorrows. If we had to choose between benefiting Smith-on-Monday and Smith-on-Tuesday, we would again face coalitional conflict. Smith-on-Monday would prefer that we benefit him, while Smith-on-Tuesday would prefer that we benefit him. Again, although typical cases are less extreme than Smith's, this clarifies the framework the Respectable Preference-Based Principle gives for reasoning about cases of proximity-bias.

The reasonableness of preferences like Jones and Smith's is not obvious or trivial. But the implications of the Respectable-Preference-Based Principle for such people is worth considering for three reasons. First, the reasonableness of such preferences is at least a matter of substantial debate: the *unreasonableness* of such preferences is neither obvious nor trivial.<sup>30</sup> Second, the plausibility of these implications gives us a means of assessing the reasonableness of future-bias and proximity-bias: if we find the Respectable-Preference-Based Principle plausible but these implications *implausible*, then we should deny the reasonableness of future-bias and proximity bias. Third, recall that others' preferences are respectable when they are *sufficiently* reasonable. Even if future-bias and proximity-bias are

---

<sup>30</sup> For a useful overview of arguments on reasonableness of future-bias, see Greene et al. 2022. For discussion of the reasonableness of proximity bias (within the context of decisions about the future), see Parfit 1984: 158–163.

unreasonable, they seem not *excessively* so: a reasonable and informed person might, in good faith, come to have such biases. Even if we cannot, on reflection, justify adopting such biases ourselves, we might have reason to respect such biases in others, at least where such biases guide our treatment of those very individuals.

Third, many of us have *fuzzy tails*, in the sense that the beginning and ends of our lives are emotionally disconnected from the rest of our lives. As a newborn, and as an older person in cases of severe cognitive decline, we can't even conceive of our day-to-day life as part of temporally-extended life. So we certainly couldn't care about the rest of our lives. In the simplest such cases, the beginning and ends of our lives then form isolated, selfish preference coalitions. If we had to choose between giving a person a benefit at age forty-five or at the very end of their life, we would again face coalitional conflict.<sup>31</sup>

To the extent that we have these patterns of concern, we depart from Ursula's unity in ways that make intrapersonal aggregation harder to justify. This seems, to me, the right result: if we instead had uniform concern for each part of our life, actions that promoted the intrapersonal sum of aggregate welfare would then be easier to justify, as such actions would be univocally preferred by each part of ourselves.

#### *4.2 Interpersonal aggregation is sometimes justified*

As far as the Respectable Preference-Based Principle is concerned, interpersonal aggregation is sometimes justifiable. This is clear from Parent & Child: both prefer that we benefit Child, even over the intrapersonally aggregative alternative of benefitting Parent.

Nor is easier-to-justify interpersonal aggregation limited to parent-child cases. Consider

---

<sup>31</sup> Intrapersonal aggregation will be less limited if the Respectable-Preference-Based Principle permits *chaining*: if my concern for my future self includes an open-ended concern for my future self's concern, and my future self cares about my far-future self, then my concern for my future self includes (unbeknownst to me) a concern for my far-future self. Chaining could also license more interpersonal aggregation. However, it's not obvious that concern for P requires such open-ended concern for P's concerns. On the face of it, I can care for my brother without caring about basketball in the same way that he does. So chaining requires more defense.

**Baseball.**

We can either

- (1) relieve Shortstop of a migraine now,
- (2) relieve Shortstop of many individually smaller but collectively greater headaches later, or
- (3) relieve many Teammates of individually smaller but collectively greater headaches,

where all teammates prefer 3 on the grounds that it will best help them win the season.

and

**Bluegrass.**

We can either

- (1) relieve Fiddler's broken finger,
- (2) relieve Fiddler's many individually smaller but collectively greater cases of tendinitis later, or
- (3) relieve many Bandmates individually smaller but collectively greater cases of tendinitis,

where all players in the band prefer 3 on the grounds that it will best help them put on a good show.

1 is implausible in each case—it's better for none and preferred by none. The Metaphysical Principle seems to favor 2 in each case, since we can justify 2 by appealing to the aggregate benefits to a single individual. But, on the Respectable-Preference-Based Principle, 3 is easier to justify. Indeed, 3 is justified on the same basic grounds as typical intrapersonal aggregation: just as a person can form a preference coalition that prefers many small benefits later to a large benefit now, a set of people can form a preference coalition that prefers many small benefits across many people to a large benefit to one person.

You might be skeptical, first, that the Metaphysical Principle really does entail that we should do 2 in each case: being on a winning team and putting on a good show seem to be goods *for* the teammates and the bandmates. To that extent, maybe the Metaphysical Principle entails that we should do 3, and the Metaphysical Principle *also* licenses something like interpersonal aggregation, if we

understand welfare capaciously enough.<sup>32</sup> However, we should not oversell this point. On many ways of spelling out the cases, Shortstop and Fiddler, like the sacrificial parent, could prefer that others benefit while appropriately conceiving of that option as a sacrifice. It must be possible to sacrifice one's own welfare to further a shared goal—to *take one for the team*.

You might be skeptical, next, that the preferences for 3 are respectable—in particular, that the Teammates' and Bandmates' preferences are reasonable. Perhaps, the thought goes, the right attitude to have toward your teammates and colleagues is *solidaristic*, such that you prefer that their burdens (and forgone benefits) be compensated by commensurate benefits *to them*, not to others or to the team or band as a whole. Then the preferences of the Teammates and Bandmates for 3 would be excessively selfish. If so, then the Respectable-Preference-Based Principle doesn't really license interpersonal aggregation in these cases, since we can ignore such unreasonable preferences. However, suppose that the Teammates' and Bandmates' preferences are *sensitive* to the sacrificiant's own preferences, such that the teammates only prefer that Shortstop sacrifice if Shortstop wants to do so, and the colleagues only prefer that Fiddler sacrifice if Fiddler wants to do so. Preferences sensitive to the sacrificiant's preferences in this way seem reasonable, not excessively selfish. The Teammates and Bandmates might even have *other-regarding reasons* to prefer that the sacrificiants sacrifice. Taking one for the team can, paradoxically, be a deep source of meaning for one's own life. The teammates and colleagues might prefer that Shortstop and Fiddler sacrifice not only to achieve a shared goal (winning the game, playing music), but also so that Shortstop and Fiddler can enrich their lives by taking one for the team.

Thus, on the Respectable-Preference-Based Principle, interpersonal aggregation can be easier to justify in cases like Baseball and Bluegrass. And the pattern of concern that justifies interpersonal aggregation in these cases may be pervasive. Think of the practices that enrich our lives, like philosophy. Bound up with my concern for philosophy and the role it plays in my life is a concern for the practice of philosophy itself. That concern is sensitive to (by partially consisting in or grounding) a preference that philosophy continue to flourish into the future. (Scheffler 2013, 2018) But philosophy's continued flourishing itself requires a kind of flourishing of philosophers: namely, their flourishing by way of achievements in philosophy. A valuable practice and its practitioners generally stand in a mutually supportive relationship. The practice's flourishing requires the flourishing of its practitioners *qua* practitioners. And the flourishing of a practitioner *qua* practitioner is sensitive to the flourishing of the practice: the value, for me, of my own activity of philosophy is bound up with my membership in a flourishing community of philosophers who can themselves take an interest in and respond to my accomplishments.

Our mutual concern for philosophy binds philosophers into a preference coalition that can license interpersonal aggregation of philosophical benefits. Just as Parent and Child might both prefer

---

<sup>32</sup> This is one way of interpreting Jeske (1993)'s proposal that interpersonal aggregation can be justified because people can be "quasi-compensated" for burdens they suffer by benefits to those with whom they are psychologically continuous.

that the child benefit, I might prefer that later philosophers receive philosophical benefits. Concretely, I might prefer not to receive a large grant if instead several later philosophers receive smaller grants, if the latter would better help philosophy as a practice to flourish. Of course, there are important differences between parents and their children and earlier and later practitioners of philosophy. For example, parents' responsibility for their children's very existence creates duties to care for them, but earlier philosophers don't obviously have an analogous duty to later philosophers. A more thorough and careful treatment of the general grounds of interpersonal aggregation licensed by the Respectable-Preference-Based Principle would be sensitive to this and other differences. But I hope these examples are enough to get the idea on the table for later refinement.

Strikingly, the Respectable-Preference-Based Principle is able to justify these forms of interpersonal aggregation *without appealing to impersonal value*. It is an ethically lightweight principle. It relies only on the idea that we should treat people in ways that they prefer. And it justifies interpersonal aggregation only in cases where the interpersonally aggregative outcome is valued *by* the affected parties. (Which is not to say that aggregation is *only* justifiable in such cases—the Preference-Based Principle only gives a sufficient condition on easier-to-justify aggregation, not a necessary condition on all-things-considered justifiable aggregation.) Accordingly, the Preference-Based Principle will be of interest to skeptics of impersonal value.<sup>33</sup> However, accepting the Preference-Based Principle does not preclude appeals to impersonal value. The Preference-Based Principle cannot explain why we should save five strangers over one stranger in the typical case. We may need to appeal to impersonal value to secure that result. But the Preference-Based Principle will still be attractive to defenders of impersonal value, since our reasons to benefit Ursula over Dave, or to benefit Child over Parent, cannot be explained by appealing to impersonal value.

#### *4.3 Interanimal aggregation is harder to justify than interpersonal aggregation*

Since it was first entertained by Nozick (1974: 39–42), the hybrid moral theory of utilitarianism for animals, deontology for people has remained a popular view in animal ethics.<sup>34</sup> The prevailing wisdom is that the differences between people and the other animals speak univocally in favor of stronger moral protections for people, such that the barriers to justifiable aggregation are stronger and more prevalent for people than for the other animals. The Respectable-Preference-Based Principle cuts against this prevailing wisdom: interanimal aggregation is in one important respect *harder* to justify than interpersonal aggregation.

---

<sup>33</sup> Like Taurek (1977) and Thomson (2008) who are skeptics of the very idea of impersonal value, or Scanlon (1998) and Wallace (2019) who are skeptics of the moral relevance of impersonal value.

<sup>34</sup> For contemporary discussions of this view Sebo 2023, Fischer et al., 2023, and other articles in *Philosophical Studies* 180 (4), a special issue on hybrid views in animal ethics. See also Kagan 2019, especially chapters eight and nine.

The crucial observation is that the other animals have narrower emotional horizons than people do. People, but not the other animals, care about a wide range of things besides themselves: traditions, practices, cultures, institutions, many other creatures, and, perhaps, humanity as a whole. Our non-selfish concerns can license interpersonal aggregation, as I argued in 4.2. But the other animals lack this wide range of concerns.<sup>35</sup> As Korsgaard says, “[E]ven if the rabbit’s life is not as important to her as yours is to you, nevertheless, for her it contains absolutely everything of value... except possibly the lives of her offspring.”<sup>36</sup> And if animals’ concerns can be respectable at all, it’s highly plausible animal selfishness, unlike human selfishness, is respectable. We could reasonably criticize a selfish human by saying, “He should know better. Can’t he see that other people matter too? He shouldn’t be indifferent to their fate.” But, given the rabbit’s capacities, it’s much less plausible that we can criticize a selfish rabbit in the same way. On the Respectable-Preference-Based Principle, then, animals lack the pattern of concern that would justify interanimal aggregation.

For example, consider

**Pigs vs People.**

We can either

- (1) let a pig lose a leg so five other pigs can keep their feet, or
- (2) let a human lose a leg so five other humans can keep their feet,

where the pig doesn’t care about the five other pigs, but the human *does* care about the five other humans and, along with the other humans, prefers 2.

Utilitarianism for animals, deontology for people would favor 1 over 2. But as far as the Respectable-Preference-Based Principle is concerned, 2 is easier to justify than 1: we can say of 2 that we are acting as the affected parties prefer, but we cannot say so of 1. We cannot justify 1 to the pig in terms of what she cares about.

In two respects, this conclusion is tentative. First, the Respectable-Preference-Based Principle is supported by the idea that actions are easier to justify if they are easier to justify to each affected party. But, even if it’s clearly important that we can justify our actions directly to individual people, it is less obviously important that we can justify our actions directly to individual animals. After all, people, but

---

<sup>35</sup> Even if social animals like ants have *some* concern for their colony that could license inter-ant aggregation, there is a large difference in scale between ants’ non-selfish concern and humans’ non-selfish concerns.

<sup>36</sup> Korsgaard 2018: 65.

not other animals, have a sense of justice.<sup>37</sup> A fuller defense of the conclusion in this subsection would require a defense of the idea that it's important that we can justify our actions directly to individual animals. Second, there are other distinctive features of humans that might count against aggregation in their case. For example, humans are autonomous. Autonomy counts in favor of *rights* for humans, rights that raise barriers to justifiable interpersonal aggregation. It's less plausible, though not uncontroversial, that the other animals are autonomous in a way that would vindicate such rights, or vindicate the extensive scheme of rights that humans plausibly have.<sup>38</sup> On balance, interanimal aggregation may be easier to justify than interpersonal aggregation. This will depend on the relative importance of these different considerations. But the Respectable-Preference-Based Principle still entails that interanimal aggregation is in one important respect harder to justify than interpersonal aggregation.

In future work, I hope to make this conclusion more firm. Indeed, I think it is the right result. The other animals have narrower emotional horizons than people do. When I die, if I've lived a good life, I can console myself with the fact that many of the things and people I care about will continue to flourish after I am gone. But the pig has no such consolation. The separateness of animals is a deeper divide than the separateness of persons.

## 5. COALITIONAL CONFLICTS

On the traditional view, the important ethical unit is the *person*, distinguished from other people by the metaphysics of people. One difficult and pressing question is whether and how we should aggregate the interests of different people. Some more specific questions include the questions of quantity, relevance, and distribution. The question of quantity is whether the *number* of people who would benefit from an action matters: do we have stronger reason to benefit more people, all else equal?<sup>39</sup> The question of relevance is whether one person's interest is always *relevant* to another: does relief of one interest always give rise to a reason that weighs against reasons to satisfy competing interests? An interest in not suffering a headache, unlike an interest in not being paralyzed, seems irrelevant to an interest in not dying. We can safely ignore the headaches, no matter how numerous.<sup>40</sup> The question of distribution is whether the distribution of interest-satisfaction across different lives matters: for instance, do we have stronger reason to prioritize the interests of those with fewer total interests satisfied?

---

<sup>37</sup> On these grounds, Rawls excludes the other animals from the realm of justice. (441–442) On the other hand, contractualists like Scanlon (1998: 177–187) and Wallace (2019: 154–155) are at least open to the idea that it is important that we can justify our actions to individual animals.

<sup>38</sup> For a useful overview, see DeGrazia 2002, especially chapter 2. But for a view on which animals should have a robust scheme of rights, see Donaldson and Kymlicka 2011.

<sup>39</sup> Taurek 1977 and Parfit 1978 offer classic competing answers to the question of quantity.

<sup>40</sup> For a helpful overview, see Horton 2021.

On the view suggested by the Respectable-Preference-Based Principle, an important ethical unit is the *preference coalition*, distinguished from other coalitions by their profile of respectable preferences. Similarly, one difficult and pressing question is whether and how we should aggregate the interests of different coalitions, in cases of coalitional conflict. Very similar questions appear under this heading. I'll introduce, but not settle, these questions.

First, there is a question of quantity: does the *size* of a coalition matter? Do we have stronger reason to meet a larger coalition's preferences, all else equal? For example, consider the following case.

**The Whim vs the One vs the Five.**

We must paint the moon. We can either

- (1) satisfy A's fleeting whim that the moon be amber,
- (2) satisfy B's stable preference that the moon be burgundy, or
- (3) satisfy C<sub>1</sub>, C<sub>2</sub>, ..., and C<sub>5</sub>'s stable preference that the moon be cerulean.

No one has a special claim that we paint the moon their color of choice. And none of the preferences are better supported by independent reasons.

Addressing this question adequately would require saying what it takes for one coalition to be larger than another, but there's an intuitive sense in which the coalition that prefers cerulean is larger than that which prefers burgundy, which is itself larger than the coalition that prefers amber. The most obvious option is to say that size matters, so we should paint the moon cerulean, or decide by a size-weighted lottery. Indeed, in politics, we think that coalitions gain strength by convincing people of their cause, and we might think the same holds for preference coalitions. But, on the other hand, the rationale for Respectable-Preference-Based Principle appeals to the importance of justification to *each* affected party, regardless of whether that party is in a coalition with many others. Perhaps we should give amber, burgundy, and cerulean equal chances.

Second, there is a question of relevance: is one coalition's preference always relevant to another's? Preferences could be made irrelevant by their strength, their content, or their grounds. For example, a weak preference for headache relief on the grounds that the relief will permit better enjoyment of short-form videos seems irrelevant to a strong preference to be cured of paralysis on the grounds that the cure would permit walking one's child down the aisle at their wedding. It is often difficult to separate the relevance of a preference that an interest be satisfied from the relevance of the interest itself. Nevertheless, the two kinds of relevance are in principle distinct, even if they turn out to be connected.

Third, there is a question of distribution: does the distribution of preference-satisfaction across different coalitions matter? If there have been a series of decisions on which preference coalition A has had their way and preference coalition B has lost out, should we give preference coalition B's preference greater weight in our current deliberation? For example, if a moose's preference for life is *all that matters* to the moose, then we might weigh it in one way more heavily than a person's preference for life, given that many of the person's *other* preferences are satisfied (e.g., the preference that the valuable practices that they care about will continue).

I won't settle these questions here. But I hope that presenting them makes clearer how the framework suggested by the Respectable-Preference-Based Principle parallels the traditional framework suggested by the Metaphysical Principle, as well as setting up avenues for further research.

## 6. WHAT JUSTIFIES AGGREGATION?

Having defended the Respectable-Preference-Based Principle, I can return to the starting question: what conditions, in general, make aggregation easier to justify? On my view, we should be pluralists about the conditions that make aggregation easier to justify. The Respectable-Preference-Based Principle offers one important condition: aggregation is easier to justify when the affected parties respectably prefer it. But the Metaphysical Principle plausibly offers another: aggregation is easier to justify when the benefited and burdened parties are the same person or animal. The Respectable-Preference-Based and Metaphysical Principles are not direct competitors. We can plausibly accept both. Indeed, there are several other plausible principles, which I will survey shortly.

But we can't answer the starting question merely by giving a list of principles. We also need to say something about the relations between the principles. To make headway on that project, I'll assess the relative importance of the Respectable-Preference-Based and Metaphysical Principles.

### 6.1 *The relative importance of the Respectable-Preference-Based and Metaphysical Principles*

We have reasons to give priority to the Respectable-Preference-Based Principle over the Metaphysical Principle when the principles conflict.

First, we have *ethical* reasons to give priority to the Respectable-Preference-Based Principle over the Metaphysical Principle. Recall Parent & Child. We have anti-paternalistic reasons to do what Parent and Child prefer, giving the benefit to Child, rather than benefiting Parent, even if benefiting Parent better respects the separateness of persons. Anti-paternalistic considerations come from, at least, the importance of autonomy as an ideal of authorship over one's own life<sup>41</sup> and the importance of relating to others as equal rational agents.<sup>42</sup> Admittedly, *preference* plays a weaker role than *choice* in justifying anti-paternalism, and anti-paternalistic considerations are less obviously weighty when we

<sup>41</sup> See, e.g., Enoch 2017, especially the discussion of non-alienation in section 8.

<sup>42</sup> See, e.g., Cholbi 2017.

think about our relations with the other animals. For other animals, the Metaphysical Principle may have greater weight.

Second, we have *methodological* reasons for thinking that the Respectable-Preference-Based Principle is in fact more fundamental than the Metaphysical Principle. The Metaphysical Principle claims that the justifiability of aggregation is conditioned on facts about personal identity. But one prominent strategy for establishing a theory of personal identity is to check whether we have reasons for certain kinds of preferences. For example, Williams (1970) aims to make progress in the theory of personal identity by asking whether, in a series of cases, an egoistic fear of future pain would be appropriate. Parfit, despite doubting the ethical significance of personal identity, takes it as a methodological constraint that we should sketch the contours of personal identity by testing the reasonableness of our strong reactions to various cases, e.g., whether we would have reasons to prefer not to go through a teletransporter. (1984: 199–200) And Shoemaker (2007) emphasizes that the appropriateness of anticipation and self-concern in various cases can guide us toward a theory of personal identity. This suggests that some normative work apparently done by persons is, in part, done by these underlying preferences.<sup>43</sup> The separateness of persons has seemed ethically significant, in part, because separate persons so often form separate preference coalitions. The plausibility of the Metaphysical Principle is in part derived from the plausibility of the Respectable-Preference-Based Principle. When persons and preference coalitions come apart, we should remember this by giving priority to the Respectable-Preference-Based Principle.

## 6.2 Other plausible principles

In addition to the Respectable-Preference-Based and Metaphysical Principles, we should plausibly also accept

*Consent Principle.* Aggregation of benefits to A and burdens to B is easier to justify when A and B consent.

On the most plausible views of consent, consent and preference come apart. I can consent to my neighbor borrowing my sugar, even if I'd prefer that she didn't borrow it. And I can prefer that someone give me a kiss, even if I haven't consented to it. Consent is an *act* (perhaps a mental act, or

---

<sup>43</sup> However, *the person* is a locus of ethical concern not only when we are satisfying preferences but when we are, say, deciding whether to keep a promise or to hold someone responsible for wrongdoing. The person is an ethical Swiss Army knife: it does work in many different parts of moral theory. We cannot conclude that preference coalitions can replace or supplement persons in other parts of morality. I see no reason to reformulate the morality of promises to make room for preference coalitions: the fact that a later person would prefer they had not made the promise does not release them from their duty to fulfill the promise. And similarly for the morality of responsibility: the fact that a later person would prefer that they had not done wrong does not release them from their duty to make amends.

perhaps an act of successful or attempted communication) while preference is a *disposition*.<sup>44</sup> On the Consent Principle, intrapersonal aggregation is not always justified. For example, consider the following case.

**Dax Cowart.**

Cowart suffered heinous injuries in a propane explosion. Cowart begged the doctors to let him die, given the severity of his pain. Despite his refusal to consent, the doctors treated him, which, as they told him would happen, let him live many decades of happy life.<sup>45</sup>

By the Consent Principle, the doctors would have more easily been able to justify their actions had Cowart consented. In general, the Consent Principle implies that I can more easily justify imposing a burden on you for the sake of benefits to you or others when you and the others validly consent to the imposition of that burden. This is a general feature of valid consent: your valid consent makes less justifiable treatment of you easier to justify. It is plausible that one can validly consent to many pointless burdens, so it is plausible that one can validly consent to a burden imposed for the sake of greater benefits. So the Consent Principle is plausible. A more thorough argument for the Consent Principle would rely on a more general account of consent.

Parfit (1984) argues for what we can call

*Early Parfit's Principle.* Aggregation of burdens to A and benefits to B is easier to justify when A and B are R-related.

where A and B are R-related, roughly, if they are psychologically continuous and connected. (Parfit 1984: 215–216) Early Parfit's Principle is most plausible as a *derivative* principle. In particular, R-relatedness often grounds people's respectable concerns for their future selves. Then, by the Respectable-Preference-Based Principle, aggregation of benefits between people and their R-related future selves will be easier to justify. But when Early Parfit's Principle conflicts with the Respectable-Preference-Based Principle, I'm inclined to say that we should give the latter priority. If I said to you, "I know you'd prefer I didn't, but I've decided to give your future, R-related self a greater benefit rather than benefit something else you respectably care about," this would be deficient as a directed justification. I would do better by you if I do what you respectably prefer.

This clarifies an advantage of the Respectable-Preference-Based Principle. When A and B stand in some relation (say, the relation of being identical, or of being R-related), that relation can *directly*

---

<sup>44</sup> See, e.g., Dougherty 2021: chapters 2 and 3 and the references there.

<sup>45</sup> See Burt and Cowart 1998 for a summary and discussion of this case.

make aggregation easier to justify. Alternatively, the relation can *indirectly* make aggregation easier to justify by bearing on the respectability of A and B's aggregation-friendly preferences. The Respectable-Preference-Based Principle then takes these respectable preferences, where they exist, as justification for aggregation. Indeed, I think this is one way of interpreting Parfit (1984)'s methodology: when we check various cases for whether relation R justifies aggregation, we are checking, most directly, whether we would ourselves respectably prefer that our burdens be balanced by benefits to those to whom we are R-related. The Respectable-Preference-Based Principle explains *why* this preference is relevant to the justifiability of aggregation.

Parfit (2011) later argues that we should act in accordance with principles that everyone could rationally choose.<sup>46</sup> This implies what we can call

*Late Parfit's Principle.* Aggregation of burdens to A and benefits to B is easier to justify when A and B could rationally permissibly choose it.

One problem with Late Parfit's Principle is that it is vulnerable to a circularity objection. Parfit justifies interpersonal aggregation on the grounds that all can rationally choose to (act in accordance with rules that) promote aggregate welfare. But you might have hoped that such a verdict would be an *output* of Parfit's theory rather than an *input*. The problem is that Late Parfit's Principle takes the principles of practical reason—the principles governing rational and justified choice—as both inputs and outputs. The principle thus constitutes limited progress in the theory of practical reason: it can tell us that interpersonal aggregation is morally justified only if we are already convinced that it is rationally permissible.

Though, as pluralists, we might accept both principles, we should grant the Respectable Preference-Based Principle priority over Late Parfit's Principle, for three reasons. First, the Respectable-Preference-Based Principle is able to avoid the circularity objection to which Late Parfit's Principle is vulnerable. Rather than taking principles of choice as both input and output, the Respectable-Preference-Based Principle takes principles of *preference* as input and outputs principles of *choice*. But different constraints apply to preference and choice. We seem to have more leeway in the realm of preference than in the realm of choice. For example, if I promised I'd order the dumplings, so you could try one, I'm required (by any plausible theory of justified choice) to order the dumplings, but it's reasonable (by any plausible theory of reasonable preference) to prefer to order the noodles. Since a preference to  $\phi$  is just a disposition to choose to  $\phi$ , choosing to  $\phi$  is one way to prefer to  $\phi$ . But there are other ways of preferring to  $\phi$ : feeling like  $\phi$ -ing, believing that there are strong reasons in favor of  $\phi$ -ing, believing that there are reasons of a certain kind (e.g., reasons grounded in one's deep

---

<sup>46</sup> Parfit 2011: 371–403.

commitments) in favor of  $\phi$ -ing. It is not immediately obvious that the same kinds of constraints apply to each way of preferring to  $\phi$ , and bringing the same kinds of requirements that apply to choice to bear on all kinds of preferences seems excessively invasive. Second, by appealing to *actual* attitudes rather than *hypothetical* attitudes (i.e., by appealing to what parties *actually* prefer, rather than what they *could* choose), the Respectable-Preference-Based Principle gives a greater role for autonomy and self-determination. Third, Parfit supports his principle by appealing to the idea that impersonal value gives everyone a reason to choose principles that maximize aggregate welfare.<sup>47</sup> But my argument for the Respectable-Preference-Based Principle nowhere appealed to this controversial idea—it appealed only to the idea that promoting aggregate welfare can hold *personal* value for the affected parties.

## 7. CONCLUSION

What conditions justify aggregation? I’ve argued that we should be pluralists. But I’ve also argued that a particular, novel condition—that the affected parties respectably prefer that we aggregate—should play a central role in our theory of justifiable aggregation. This Respectable-Preference-Based Principle implies that intrapersonal aggregation is not always maximally easy to justify, that interpersonal aggregation is sometimes easier to justify than intrapersonal aggregation (without appealing to impersonal value!), and that, against conventional wisdom, interanimal aggregation is in one important respect *harder* to justify than interpersonal aggregation.

I’ve left many important questions open, and left others with only tentative answers, including the questions of what to do in cases of coalitional conflict and questions of relative importance and fundamentality of the various conditions that justify aggregation.<sup>48</sup> The success of this project will ultimately depend on whether we can give satisfactory answers to these further questions. But I hope my arguments here suffice to establish the defensibility of a new way of thinking about aggregation.

## References

- Anscombe, G.E.M. 1967. “Who is Wronged? Philippa Foot and ‘Double Effect’,” *The Oxford Review* 5 (1967), 16-17.
- Barry, B. 1991. “Justice Between Generations,” in *Liberty and Justice: Essays in Political Theory* 2, pp. 242-258.
- Bourguignon, Léa. 2024. “On the Possibility of Act Contractualism.” *Australasian Journal of Philosophy* 103 (3):756–774.

---

<sup>47</sup> Parfit 2011: 371–403.

<sup>48</sup> I also haven’t discussed complications that arise when we introduce risk. We might want to know not only when we can justifiably aggregate burdens to A and benefits to B, but when we can justify imposing *risks* of burdens and benefits on A and B in ways that appeal to (expected or actual) aggregate burdens and benefits. These questions are important, but I lack the space to give them the treatment they deserve. For some helpful discussions of these issues, see Buchak (2017), Frick (2015), Otsuka and Voorhoeve (2009), Voorhoeve and Fleurbaey (2012), and Van Fossen (2024).

- Brink, D. 1993. "The Separateness of Persons, Distributive Norms, and Moral Theory." In Raymond Gillespie Frey & Christopher W. Morris, *Value, Welfare, and Morality*. Cambridge University Press. pp. 252-289.
- Buchak, L. 2017. "Taking Risks Behind the Veil of Ignorance." *Ethics* 127 (3):610-644.
- Burt, R. & Cowart, D. 1998. "Confronting Death Who Chooses, Who Controls?" *Hastings Center Report* 28 (1):14-24.
- Chang, R. 2009. "Voluntarist reasons and the sources of normativity." In David Sobel & Steven Wall, *Reasons for Action*. New York: Cambridge University Press. pp. 243-71.
- Cholbi, M. 2017. "Paternalism and our Rational Powers." *Mind* 126 (501):123-153.
- Darwall, S. 2001. "Because I Want It." *Social Philosophy and Policy* 18 (2):129-153.
- 2002. *Welfare and Rational Care*. Princeton University Press.
- DeGrazia, D. 2002. "The moral status of animals," *Animal Rights: A Very Short Introduction, Very Short Introductions* (Oxford, 2002; online edn, Oxford Academic, 24 Sept. 2013), <https://doi.org/10.1093/actrade/9780192853608.003.0002>
- Doggett, T. 2011. "Saving the Few." *Noûs* 47 (2):302-315.
- Donaldson, S. & Kymlicka, W. 2011. *Zoopolis: A Political Theory of Animal Rights*. Oxford.
- Dougherty, T. 2021. *The Scope of Consent*. Oxford.
- Enoch, D. 2017. "Hypothetical Consent and the Value (s) of Autonomy." *Ethics* 128 (1):6-36.
- Foot, P. 1983. "Utilitarianism and the Virtues." *Proceedings and Addresses of the American Philosophical Association* 57 (2):273-283.
- Fischer, B.; Palmer, C.; & Kasperbauer, T. 2023. "Hybrid theories, psychological plausibility, and the human/animal divide." *Philosophical Studies* 180 (4):1105-1123.
- Fischer, J. 2021. "Counting People and Making People Count." *Philosophy* 96 (2):229-252.
- forthcoming. "Consequentialism and the separateness of persons." *Philosophical Quarterly*.
- Frankfurt, H. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1):5-20.
- Frick, J. 2015. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43 (3):175-223.
- Frick, J. 2016. "What We Owe to Hypocrites: Contractualism and the Speaker-Relativity of Justification." *Philosophy and Public Affairs* 44 (4):223-265.
- Gauthier, D. 1963. *Practical Reasoning*. Oxford: Clarendon Press.
- Greene, P.; Latham, A.; Miller, K.; Norton, J.; Tarsney, C. & Tierney, H. 2022. "Bias Towards the Future." *Philosophy Compass* 17 (8):1-11.
- Halstead, J. 2016. "The Numbers Always Count." *Ethics* 126 (3):789-802.
- Heathwood, C. 2017. "Which Desires Are Relevant to Well-Being?" *Noûs* 53 (3):664-688.
- Hirose, I. 2013. "Aggregation and the Separateness of Persons." *Utilitas*. 25 (2): 182-205.
- Hirose, I. 2014. *Moral Aggregation*. New York, US: OUP USA.

- Horton, J. 2018. "Always Aggregate." *Philosophy and Public Affairs* 46 (2): 160-174.
- 2021. "Partial aggregation in ethics." *Philosophy Compass* 16 (3):1-12.
- Hume, D. 1738/2000. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Oxford, GB: Oxford University Press. Edited by David Fate Norton & Mary J. Norton.
- Hurka, T. 2019. "More Seriously Wrong, More Importantly Right." *Journal of the American Philosophical Association* 5 (1): 41–58. <https://doi.org/10.1017/apa.2018.41>.
- Jeske, D. 1993. "Persons, Compensation, and Utilitarianism." *Philosophical Review* 102 (4):541-575.
- Kamm, F. 1993, *Morality Mortality. Volume. I: Death and Whom to Save From It*. Oxford University Press.
- 2007, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford University Press.
- Kolodny, N. 2005. "Why Be Rational?" *Mind* 114 (455):509-563.
- Korsgaard, C. 1996. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge University Press.
- Korsgaard, C. 2013. "The Relational Nature of the Good." In Russ Shafer-Landau, *Oxford Studies in Metaethics*, Volume 8. Oxford University Press. pp. 1-26.
- Korsgaard, C. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press.
- Kumar, R. 2001. "Contractualism on saving the many." *Analysis* 61 (2):165–170
- Lazar, S. 2018. "Limited Aggregation and Risk." *Philosophy and Public Affairs* 46 (2):117-159.
- McIntosh, R. ms. "Doing What Another Would Want."
- McMahan, J. 2002. *The Ethics of Killing: Problems at the Margins of Life*. OUP Usa.
- Munoz-Dardé, V. 2005. "The Distribution of Numbers and the Comprehensiveness of Reasons." *Proceedings of the Aristotelian Society* 105 (2):207–233.
- Nagel, T. 1972. "War and Massacre." *Philosophy & Public Affairs* 1(2):123–144.  
<http://www.jstor.org/stable/2264967>
- Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Otsuka, M. 2006. "Saving lives, moral theory, and the claims of individuals." *Philosophy and Public Affairs* 34 (2):109–135.
- Otsuka, M. & Voorhoeve, A. 2009. "Why it matters that some are worse off than others: An argument against the priority view." *Philosophy and Public Affairs* 37 (2):171-199.
- Parfit, D. 1973. "Later Selves and Moral Principles" in Alan Montefiore, *Philosophy and personal relations*. Montreal,: McGill- Queen's University Press. pp. 137-169
- Parfit, D. 1978. "Innumerate Ethics." *Philosophy & Public Affairs* 7 (4):285-301.
- 1984. *Reasons and Persons*. Oxford, GB: Oxford University Press.
- Rawls, J. 1971/1999. *A Theory of Justice: Revised Edition*. Cambridge, MA: Harvard University Press.

- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge: Belknap Press of Harvard University Press.
- Scheffler, S. 2013. *Death and the Afterlife*. New York, NY: Oup Usa. Edited by Niko Kolodny.
- Scheffler, S. 2018. *Why Worry About Future Generations?* Oxford University Press.
- Sebo, J. 2022. "Kantianism for humans, utilitarianism for nonhumans? Yes and no." *Philosophical Studies* 180 (4):1211-1230.
- Shoemaker, D. 2007. "Personal Identity and Practical Concerns." *Mind*. 116 (462):317-357.
- Setiya, K. 2014. "Love and the Value of a Life." *Philosophical Review* 123 (3):251-280.
- Sidgwick, H. 1874/1981. *The Methods of Ethics*. Indianapolis, IN: Hackett.
- Steuwer, B. 2020. *One-by one: moral theory for separate persons*. Dissertation. London School of Economics.
- Street, S. 2009. "In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters." *Philosophical Issues* 19 (1):273-298.
- Taurek, J. 1977. 'Should the Numbers Count?' *Philosophy and Public Affairs* 6 (4):293-316.
- Thomson, J.J. 2008. *Normativity*. Open Court. Edited by Russ Shafer-Landau.
- Temkin, L. 2011. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. US: Oxford University Press.
- Van Fossen, S. 2024. "Can Relative Prioritarianism Accommodate the Shift?" *Ethics* 134 (4):525-538.
- Voorhoeve, A. 2014. "How Should We Aggregate Competing Claims?" *Ethics* 125 (1):64-87.
- Voorhoeve, A. & Fleurbaey, M. 2012. "Egalitarianism and the Separateness of Persons." *Utilitas* 24 (3):381-398.
- Wallace, R.J. 2019. *The Moral Nexus*. Princeton University Press.
- Watson, G. 1975. "Free Agency." *Journal of Philosophy* 72 (8):205-20.
- Yetter Chappell, R. 2015. "Value Receptacles." *Noûs* 49 (2):322-332.
- Zhang, E. 2024. "Individualist Theories and Interpersonal Aggregation." *Ethics* 134 (4):479-511.