Workshop on Welfare and Ethics

27–28 June 2024, Saarland University

Thursday, 27 June

11:00 – 11:15	*Welcome and Introduction*
11:15 – 12:00	Jonas Harney (Saarland University) *Welfare Losses in Variable Population Choices with Multiple Options*
12:30 – 13:15	Tomi Francis (Oxford University) *An Intrapersonal Argument for Totalism*
	Lunchbreak
14:15 – 15:00	Adriano Mannino (UC Berkeley) *Welfarist Population Ethics and the Future of Humanity*
15:30 – 16:15	Jakob Lohmar (Oxford University) *Against Probability-Discounting when Testing for Relevance*
16:45 – 18:15	Joe Horton (University College London) *to be announced...*
19:30	*Dinner*

Friday, 28 June

10:00 – 10:45	Abelard Podgorski (National University of Singapore) *Better and Worse Ways Never to Exist*
11:15 – 12:00	Alexander Sasha Arridge (Oxford University) *What We May Claim for Ourselves*
12:30 – 13:15	Martin Dimitrov (University of Toronto) *The Centrality of Pleasure to Well-being*
	Lunchbreak
14:15 – 15:00	Eva Rose Read (London School of Economics) *Animal Welfare as Holistic Self-Maintenance: Against the Valence Aggregation Model*
15:30 – 17:00	Theron Pummer (University of St. Andrews) *Hypersensitivity: No Vague Escape*
19:00	*Dinner*

Jonas Harney (Saarland University)
Welfare Losses in Variable-Population Choices with Multiple Options

We've moral reasons not to make people worse off than they could have been. Call this the

> *Comparative View*
> The extent to which an option A would be worse for an individual than another option
> B provides in itself a moral reason against A.

The Comparative View is deeply plausible. Yet, it can be argued to imply cyclical obligations as the following case shows, where the numbers represent the individuals' welfare level and "---" denotes that an individual does not exist.

	Ali	**Bel**	**Cam**
A	100	50	----
B	----	100	50
C	50	----	100

We have loss-provided moral reasons to choose B over A due to Bel's loss, to choose C over B due to Cam's loss, and to choose A over C due to Ali's loss. Given that everything else is equal in the three outcomes and assuming that we ought to choose what we have strongest moral reason to choose, we also have moral obligations to choose B over A, C over B, and A over C. Hence, whatever we choose, there is always an option that we ought to choose instead. This is implausible.

However, the objection implicitly assumes that a welfare loss in an option *A* relative to another option *B* provides only a *pairwise* reason to choose *A* over *B*. If the loss-provided reasons are *set-wise*, by contrast, a loss in *A* relative to *B* provides reason to choose any available option over *A*. If so, we have equally strong loss-provided reasons against A, B, and C because, in each of the options, one person loses 50 units of welfare. On the set-wise understanding, therefore, the welfare losses do not provide a reason to choose any of the three options over another. Thus, the set-wise understanding of the loss-provided reasons avoids cyclical obligations.

Alternatively, we could supplement the pairwise understanding with a tournament decision-theory (Cf. Ross 2015, Podgorski 2023). The loss-provided reasons are essentially pairwise, but they don't imply obligations. Rather, they give rise to defeat-relations between the options. Applying decision-theoretical rules then allows us to avoid cyclical obligations. However, the tournament model is insensitive to different welfare losses and, thus, different strengths of the reasons provided by these welfare losses.

The set-wise understanding implies seemingly implausible implications known as *improvable live avoidance* (Ross 2015) or *backfiring complaints* (Horton 2021).

	Ali	Bel
A	10	----
B	10	10
C	–1000	1000

On the set-wise understanding, we ought to choose B over A because Bel's loss in B relative to C provides a strong reason against B and no reason counts against A. While this seems initially implausible, I argue that we should accept it. The explanations for why the implication would be implausible given by Ross (2015), Horton (2021), and Podgorski (2023) are unwarranted, and so is the attempt to explain it by an appeal to a moral difference between the welfare losses of individuals who exist dependently on and those who existing independently from the choice. If we take welfare losses seriously, the implications are perfectly plausible.

References

Horton, Joe. 2021. 'New and Improvable Lives'. *The Journal of Philosophy* 118 (9): 486–503.

Podgorski, Abelard. 2023. 'Complaints and Tournament Population Ethics'. *Philosophy and Phenomenological Research* 106 (2): 344–367.

Ross, Jacob. 2015. 'Rethinking the Person-Affecting Principle'. *Journal of Moral Philosophy* 12 (4): 428–461.

27 June, 12:30 – 13:15

Tomi Francis (Oxford University)

An Intrapersonal Argument for Totalism

This paper provides an argument for Totalism: the view that one population is better than another iff it contains greater total wellbeing. It is an intrapersonal argument: an argument that appeals to pareto-like principles which say that one option is better than another if it makes everyone better off. It applies such principles to cases of risk (as has been done for decades), and to cases in which some of the individuals have a chance of existence, but are not guaranteed to exist (a more recent development pioneered by McCarthy, Mikkola, and Thomas 2020 and Nebel 2019, among others). The argument has two parts. First I argue for Risky Existence Totalism: the view that prospect X is better for Alice than prospect Y if and only if X gives Alice a greater expected level of wellbeing, given that states of nature in which Alice does not exist make no difference to this expected level of wellbeing. Next, I argue from Risky Existence Totalism to Totalism proper.

The first part of the argument goes along the following lines. If we can ever compare, in terms of individual betterness, two prospects which give Alice different probabilities of existence, then we can do so in the following sort of case: prospect X guarantees that if Alice exists, she has a good life, while prospect Y guarantees that if Alice exists, she has a bad life. In such cases, prospect X is better for Alice than prospect Y. Call this the Conditional Value Principle.

Now imagine that a number between one and three will be drawn randomly, with equal probabilities. In prospect X, Alice is guaranteed to exist, and has twenty units of wellbeing. In prospect Y, Alice exists if the numbers one or two are drawn, and has twenty five units of wellbeing

in either case. If the number three is drawn, Alice does not exist. We can represent Alice's prospects like this:

	One is drawn	Two is drawn	Three is drawn
X	20	20	20
Y	25	25	–

Now consider two more prospects. Rather than describe them, I'll just give you the table:

	One is drawn	Two is drawn	Three is drawn
X′	54	3	3
Y′	54	−4	–

Prospect X' gives Alice the same expected wellbeing as prospect X. Similarly, prospect Y' gives Alice the same expected wellbeing as prospect Y . Thus, X and X' are equally good for Alice; similarly, Y and Y' are equally good for Alice.[1] Next, consider another two prospects:

	One is drawn	Two is drawn	Three is drawn
X″	–	3	3
Y″	–	−4	–

These two prospects differ from X' and Y' only in that Alice fails to exist, rather than getting 54 units of wellbeing, if the number one is drawn. But the choice between X'' (or X') and Y'' (or Y') surely cannot depend on what happens if the number one is drawn, because these choices make no difference to what happens if the number one is drawn.[2] We should therefore say the same thing about X' and Y' as we say about X'' and Y''. The Conditional Value Principle implies that X'' is better for Alice than Y''. X' is therefore better for Alice than Y'. Since X' is equally as good as X, and Y' is equally as good as Y , we can also say that X is better for Alice than Y , in line with Risky Existence Totalism.

The argument above can be generalised: it turns out that Risky Existence Totalism follows from the Conditional Value Principle, the claim that we should maximise expected wellbeing when we are sure to exist, and the Sure Thing Principle. And it turns out that Totalism follows from Risky Existence Totalism, given a version of Anonymity, a principle of Statewise Dominance, and an Ex

---

[1]   I am assuming here that the scale on which Alice's wellbeing is being measured is one which accords with a utility function representing the betterness relation for Alice for prospects where she is guaranteed to exist. Such a utility function exists if this betterness relation satisfies the axioms of Expected Utility Theory, the details of which need not bother us here. The key point is that Alice's wellbeing on this scale may not accord with our intuitive conception of "the amount of wellbeing" at stake for Alice. So my argument in this paper is really an argument for the version of Totalism where wellbeing is measured in a way which accords with whichever degree of risk aversion it is rational to have, rather than the version of Totalism where wellbeing is measured independently of considerations of risk.

[2]   This is the Sure Thing Principle; see Savage 1954.

Ante Pareto principle of the sort I mentioned earlier. I don't have the space to explain this second part of the argument here, but basically it's a version of Harsanyi's Aggregation Theorem.[3]

References

Broome, John (1991). Weighing goods: equality, uncertainty and time. Oxford: Basil Black- well.

Fleurbaey, Marc (2009). "Two Variants of Harsanyi's Aggregation Theorem". In: Economics Letters 105.3, pp. 300– 302.

Harsanyi, John C. (1955). "Cardinal Welfare, Individualistic Ethics, and Interpersonal Com- parisons of Utility". In: Journal of Political Economy 63.4, pp. 309–321.

McCarthy, David, Kalle Mikkola, and Teruji Thomas (2020). "Utilitarianism With and Without Expected Utility". In: Journal of Mathematical Economics 87, pp. 77–113. Nebel, Jacob M. (2019). "An Intrapersonal Addition Paradox". In: Ethics 129.2, pp. 309– 343.

Savage, Leonard J. (1954). The Foundations of Statistics. New York: John Wiley & Sons.

27 June, 14:15 – 15:00

Adriano Mannino (UC Berkeley)
Welfarist Population Ethics and the Future of Humanity

Most contemporary work in population ethics operates within the framework of welfarism, i.e., the assumption that individual welfare is the fundamental value. Further common assumptions include impartiality and interpersonal aggregation. I argue that these commitments generate intuitively troubling and theoretically un- compelling implications about how we ought to shape the future of humanity. If we wish to avoid these implications, we must restrict the validity of welfarism and impartiality. Interpersonal aggregation may be defensible, but a broader commitment to reasons aggregation or reasons "pro-tantism" is incompatible with the specific way in which impartial welfarism must be restricted.

I start by discussing the welfarist case for longtermism. Roughly speaking, longtermism is the view that we should prioritize improving humanity's long-term future above all else (cf. Greaves & MacAskill 2019, Beck- stead 2013, and Bostrom 2003). The case for this view is based on three main premises: (i) The long-term future of humanity contains sufficiently large aggregate welfare stakes, (ii) symmetry in population ethics, i.e., we have sufficiently strong reason not just to avoid creating miserable lives, but to bring happy lives into existence, too, and (iii) we are not clueless about how to suitably and sufficiently affect the long-term future. Normative critics of longtermism often reject (ii) and claim that if we endorse the asymmetry in population ethics (as they argue we should), we can block the case for longtermism and other revisionary implications (cf., e.g., Setiya 2022 and Frick 2020).

Drawing on Broome (2005), Frick (2020) holds that asymmetric views in population ethics, in contrast to symmetric ones, do not face problems such as the following: When we save people's lives, by making roads safer or in other ways, the welfare of the people who are saved is usually small in comparison to the welfare of all the new people (their descendants) who come into existence as a

---

[3]   See Harsanyi 1955, Broome 1991, Fleurbaey 2009 and McCarthy, Mikkola, and Thomas 2020.

result. If we counted the descendants' welfare too – as impartial welfarism recommends we do –, that would enormously alter the value we intuitively attach to saving people's lives. (Public policy routinely ignores the descendants' welfare.) However, pace Frick, it is not the case that the asymmetry allows us to avoid highly revisionary implications here: If the set of descendants is sufficiently large, it will contain many miserable lives, too.[4] Therefore, asymmetric welfarist views will usually militate against saving people: If we count the miserable descendants too, their negative welfare will greatly outweigh the welfare stakes for the people saved (or let die) in the present.

By the same token, asymmetric views militate in favor of what I call *asymmetric longtermism*, the view that minimizing the number of miserable lives in our long-term future matters above all else. In practice, this view will likely recommend that we try and keep our future population size as small as possible. I argue that asymmetric views such as Frick's (2020), Kolodny's (2022), and Bader's (2022) are committed to asymmetric longtermism, at least prima facie. Frick has attempted to block troubling implications like antinatalism and extinctionism, which may follow from the general recommendation to minimize our future population size (Frick 2017 and 2014, ch. 3). However, even if we grant Frick's points – e.g., that the human civilization has final non-welfarist value and there is strong pro tanto reason to preserve it –, Frick cannot robustly avoid asymmetric longtermism. This is because, given a sufficient number of expected future generations (conditional on non-extinction in the present), the number of miserable lives in our future can become arbitrarily large. Accordingly, the pro tanto reason to minimize these miserable lives can acquire arbitrarily great strength. Hence, if this reason recommends extinction, it will outweigh the pro tanto reason to preserve human civilization.[5]

It is important to note that symmetric welfarist views do not robustly allow us to prioritize the survival of humanity either. They imply, e.g., that it would be preferable to create sufficiently many nonhuman welfare subjects de novo (on Earth, to replace humanity, or elsewhere in the universe) over preventing human extinction, provided that the newly created nonhumans have positive welfare on average.[6]

---

[4] MacAskill (2022, ch. 9, fn. 27) cites an unpublished psychological study (Caviola et al. 2021), commissioned by MacAskill himself, which found that more than 10 percent of Americans and Indians assess their own lives as containing more suffering than happiness. Even if the true figure were much smaller than 10 percent, the point would stand.

[5] In conversation, Frick has suggested that if the strength of the pro tanto reason to preserve humanity is a function of how long our future would be (conditional on non-extinction in the present), then it could avoid being outweighed by the pro tanto reason to minimize the vast number of miserable lives our long future would likely contain. There are two problems with this suggestion. First, while the strength of the reason to minimize miserable lives grows linearly in the number of miserable lives (on grounds of impartiality), preserving human civilization for two million years is arguably not twice as important as preserving it for one million years, all else equal. Second, if it were twice as important and the respective function were linear, this would turn Frick into a *humanity preservation longtermist*: Given a sufficiently long future (conditional on non-extinction in the present), the pro tanto reason to preserve humanity would overwhelm all other concerns. This is not what Frick wishes to maintain, and it is indeed deeply implausible.

[6] I argue that views such as Horton's (2021), which maintain that new happy lives have cancelling or o6setting force, face similar problems.

What kind of view, then, would allow us to avoid the troubling implication, generated by impartial welfarism, that we may have to accept human extinction or enormous hardship in the present to optimize the long-term future? I argue that such a view would have to include (i) a strict (not merely pro tanto)[7] non-welfarist commitment to the preservation of the human species, community, and civilization, and (ii) a strict commitment to near-generational partiality, to wit, strict partiality toward the present generation, including our children and grandchildren. If these commitments were merely pro tanto, they could (and likely would) be overwhelmed by the pro tanto reason to minimize miserable lives and/or the pro tanto reason to maximize happy lives, which both militate in favor of longtermism. I suggest a way to operationalize (i) and (ii) that avoids fanaticism about these commitments under uncertainty (weak lexicality); I draw on Scheffler (2013), Frick (2017), Knutzen (2023), and Masny (ms.) to tentatively justify the commitments' content; and I propose an argument from legitimate self-interest to justify the commitments' *strict* nature: Arguably, I am not required to accept sufficiently bad personal outcomes (e.g., prolonged torture or highly premature death) to improve the long-term future. I may strictly commit to avoiding such outcomes, and perhaps I should. Similarly, we may strictly commit to our collective survival and to privileging our own generation, and perhaps we should. If we do so, the practical upshot is this: Once we have lived up to these commitments to a sufficient (non-fanatical) degree, all excess resources can and arguably should go to longtermist projects, on impartial welfarist grounds.

References

Bader, Ralf. 2022. "The Asymmetry." In: McMahan, Jeff. et al. (eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford University Press.

Beckstead, Nick. 2013. *On the Overwhelming Importance of Shaping the Far Future*. Dissertation, Rutgers University.

Bostrom, Nick. 2003. "Astronomical Waste: The Opportunity Cost of Delayed Technological Development." *Utilitas* 15(3): 308–314.

Broome, John. "Should We Value Population?". *The Journal of Political Philosophy* 13(4): 399–413. Frick, Johann. 2020. "Conditional Reasons and the Procreation Asymmetry." *Philosophical Perspectives* 34(1): 53–87.

Frick, Johann. 2017. "On the Survival of Humanity." *Canadian Journal of Philosophy* 47(2-3): 344–367.

Frick, Johann. 2014. *'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics*. Dissertation, Harvard University.

Greaves, Hilary, and William MacAskill. 2019. "The Case for Strong Longermism." *GPI Working Papers*.

Horton, Joe. 2021. "New and Improvable Lives." *Journal of Philosophy* 118(9): 486–503.

Knutzen, Jonathan. 2023. "Unfinished Business." *Philosophers' Imprint* 23(1): 4, 1–15 (2023).

Kolodny, Niko. 2022. "Saving Posterity From a Worse Fate." In: McMahan, Jeff, et al. (eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford University Press.

MacAskill, William. 2022. *What We Owe the Future*. Basic Books.

Masny, Michal. ms. "The Shape of History."

Scheffler, Samuel. 2013. *Death and the Afterlife*. Oxford University Press.

Setiya, Kieran. 2022. "The New Moral Mathematics." *Boston Review*.

Taurek, John. 1977. "Should the Numbers Count?" *Philosophy and Public Affairs* 6(4): 293–316.

Vallentyne, Peter, and Shelly Kagan. 1997. "Infinite Value and Finitely Additive Value Theory." *Journal of Philosophy*

---

[7] As announced at the outset, reasons aggregation or reasons "pro-tantism" fails. This is where it does so.

27 June, 15:30 – 16:15

Jakob Lohmar (Oxford University)
Against Probability-Discounting when Testing for Relevance

According to *relevance views,* moral factors (such as benefits, harms, claims, etc.) that are generally morally relevant can be irrelevant in certain decision situations. For example, minor benefits such as the alleviation of headaches are irrelevant, on such views, when major benefits such as the saving of a life are at stake.[8] What seems to determine whether a given benefit or harm is relevant in a given decision situation is, at least among others, how large its magnitude is relative to the magnitude of the largest benefit or harm that is at stake in this decision situation.[9] A natural extension of this way of testing for relevance to decisions under *risk* is to compare the *probability-weighted* magnitudes of the benefits and harms that are at stake. Indeed, standard relevance views imply some form of probability-discounting: e*x ante* views imply that benefits and harms need to be weighted by the probability that *particular individuals* will receive them, and common *ex post* views imply that benefits and harms need to be weighted by the probability that *someone* will receive them.[10] While much of the debate on "partial aggregation" has focused on the question whether ex ante or ex post views are preferable, I will discuss the question whether *any* form of probability-discounting – whether of the ex ante or the ex post type – should be employed when testing for relevance. This question is of major practical importance since some actions that do a lot of good in expectation, such as actions that slightly decrease the risk of catastrophic events, benefit *no one* with a high probability. *Any* form of probability-discounting therefore tends to imply that the benefits of such actions are irrelevant when safe benefits could be produced instead.

In this paper, I argue that no form of probability-discounting should be employed when testing for relevance; whether a given benefit or harm is relevant is independent of all probabilities associated with it. My argument for this claim can be summarized as follows: each way of probability-discounting suffers from the *statistical lives problem* or a particularly damaging version of the *expected lives problem*. The well-known *statistical lives problem* is the implication of ex ante views that, roughly, one life of an identified person can outweigh any number of lives of unidentified people or "statistical lives". Against what has been suggested in the literature (Frick 2015), I argue that the statistical lives problem cannot be sufficiently resolved by *pluralist* ex ante views either. All forms of probability-discounting that avoid the statistical lives problem (which correspond to a subclass of all ex post views) suffer from the often overseen expected lives problem. This is the implication that,

---

[8]   Relevance views are an important subclass of partially aggregative views. See Scanlon (1998) for the general idea of relevance and e.g. Voorhoeve (2014) for a fully developed relevance view.

[9]   This criterion for relevance could be refined in several ways that do not affect my arguments in this paper. For example, benefits to the worse-off could be given extra weight (see Voorhoeve 2014).

[10]   Advocates of ex ante views include Frick (2015), James (2012), Kumar (2015), and Steuwer (2021). Advocates of ex post views include Otsuka (2015), Reibetanz (1998), and Rüger (2018).

roughly, one statistical life can outweigh any expected number of lives. While this basic version of the problem is not decisive by itself, it can be exacerbated in several ways which I demonstrate with decision tables.

While the focus of my paper is on arguing against probability-discounting, I also sketch an alternative way of extending relevance views to decisions under risk that does not suffer from any of the discussed problems.

References

Frick, Johann. (2015). Contractualism and Social Risk, Philosophy & Public Affairs, 43(3): 175-223.

James, Aaron (2012), "Contractualism's (Not So) Slippery Slope," Legal Theory 18: 263–92.

Kumar, Rahul (2015), "Risking and Wronging," Philosophy & Public Affairs 43: 27–49.

Otsuka, Michael. (2015). Risking life and limb, in I. Glenn Cohen, Norman Daniels, and Nir Eyal (eds). Identified versus Statistical Lives: An Interdisciplinary Perspective, Oxford: Oxford University Press: 77-93.

Reibetanz, Sophia. (1998). Contractualism and Aggregation, Ethics, 108(2): 296-311.

Rüger, Korbinian. (2018). On ex ante contractualism. J. Ethics & Soc. Phil. 13: 240.

Scanlon, T. M. (1998). What We Owe to Each Other, Cambridge, MA: Belknap Press of Harvard University Press.

Steuwer, Bastian. (2021) Contractualism, Complaints, and Risk, Journal of Ethics and Social Philosophy, 19(2): 111-147.

Voorhoeve, Alex. (2014). How Should We Aggregate Competing Claims? Ethics, 125(1): 64-87.

27 June, 16:45 – 18:15

Joe Horton (University College London)
to be announced...

June 28, 10:00 – 10:45

Abelard Podgorski (National University of Singapore)
Better and Worse Ways Never to Exist

In this talk, I defend the absurd-seeming view that some ways of never existing are better for the person who never exists than others, and discuss its normative upshot. Roughly, the strategy is as follows: first, I consider a much more common, though still controversial view – that someone's existence can be better or worse for them than their nonexistence. I will briefly discuss why this view is attractive, and introduce a well-known objection – that it is incoherent to claim that nonexistence is better or worse for someone than existence, because someone must exist to have well-being.

I suggest this objection fails because it conflates the question of whether a world is better or worse for a person than another with the question of whether that person's life is better or worse in that world. It is clearly absurd to say that my existing life is better than the life I would have had if I didn't exist. But as Wlodek Rabinowicz and others have pointed out, this is not the only way to understand the relevant comparative claims, and we can maintain logical consistency provided that we accept that claims about welfare comparisons between worlds are *contingent*.

Whether for the reason I propose or for other reasons discussed in the literature, if we accept that there is no logical problem comparing the value for me of a world in which I don't exist with the value for me of a world in which I do, then there is also no logical problem comparing the value for me of two worlds in which I don't exist. Whether any two such worlds are equally good for me is then a question for a substantive theory of welfare.

On some theories of welfare, such as hedonism, it will turn out that all worlds empty of me are equally good for me (since I have no pleasure in any of them). But I argue that a plausible version of desire-satisfactionism, which claims that how good a world is for me depends on how well it matches my desires in the actual world, implies that some ways of not existing are better for me than others. For example, if I have an important life project, such as preserving the beauty of nature, then a world in which I never exist but nature is preserved is better for me than a world in which I never exist and nature is destroyed.

For the remainder of the talk, I will respond to some potential objections to this substantive view of welfare, such as that only certain kinds of desires can matter for welfare, and that the reference to the actual world makes what I ought to do objectionably depend on what I will do. Finally, I consider what normative relevance these comparisons between nonexistences could possibly have, proposing that they matter for the fittingness of certain attitudes like regret.

<div align="right">June 28, 11:15 – 12:00</div>

Alexander Sasha Arridge (Oxford University)
What We May Claim for Ourselves

As the autonomous subjects of our wellbeing, we each have the normative power to determine how much our wellbeing matters; this power consists in the ability to modify the strength of reasons grounded in the constituents of our wellbeing. The moral limits within which we may permissibly exercise this normative power are determined by our directed duties: what we owe constrains what we may claim for ourselves. The existence of this morally bounded normative power to determine how much our wellbeing matters explains, amongst other things, the possibility of supererogation: when our directed duties allow, we are morally permitted to modify the weight of our prudential reasons so as to give ourselves most reason overall to do an otherwise morally suboptimal act. These, anyway, are the central claims of this paper.

The paper starts by arguing that we have the normative power to modify the weight of reasons grounded in the constituents of our wellbeing (e.g. our prudential reasons and others' welfarist reasons with respect to us) because we are autonomous with respect to our wellbeing. A nun, for example, makes the pleasures of sex less strongly reason-giving for her by committing to a life of celibacy: the nun's autonomous choice to be celibate normatively determines how much this constituent of her wellbeing matters to her. The nun's choice *also* modifies the weight of *our* welfarist reasons with respect to this constituent of her wellbeing: out of respect for the nun's autonomy, we

must allow her choices to determine the strength of our welfarist reasons with respect to her. As the autonomous subject of her wellbeing, the nun controls how much her wellbeing matters.

The *moral limits* within which we may permissibly exercise our normative power are, I argue, determined by our *directed duties*. Our directed duties to ourselves set a *lower* limit on how reason-giving we may make our wellbeing: we owe it to ourselves not to *entirely* discount our wellbeing. More importantly, our directed duties *to others* (e.g. promissory obligations and duties to respect rights) set an *upper* limit on how strongly reason-giving we can make our wellbeing: a parent's duty to care for their child, for example, sets an upper limit on how reason-giving they may make their wellbeing. On this view, our directed duties don't just give us moral reasons, but they also set limits on the possible strength of our prudential reasons; that our directed duties do this captures the sense in which these duties have "pre-emptive" force over our prudential reasons. The combination of our normative power and the constraints of directed duty yield a *bounded moral prerogative* to determine how much our wellbeing matters.

The paper closes by claiming that recognising the existence of this prerogative allows us to explain the possibility of supererogation and to resolve, from the ground up, Joe Horton's infamous "All or Nothing" problem. When our directed duties allow, we are morally permitted to modify the weight of our prudential reasons so as to give ourselves most reason overall to do an otherwise morally suboptimal act: outside of what we owe, we are morally entitled to do morally sub-optimal acts.

June 28, 12:30 – 13:15

Martin Dimitrov (University of Toronto)
The Centrality of Pleasure to Well-being

The main objections to hedonism convincingly show that (partly) mind-independent states of affairs like achievement contribute to well-being. However, they do not show that these states independently contribute to well-being because if this was the case, then an instance of knowledge or achievement can be really good for us though we do not care about it and derive no joy from it. Since this is implausible, there seems to be a resonance constraint on well-being: "what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive…" (Railton, 1986: 9). Though these remarks are suggestive, what exactly does it mean for something like one's achievement or knowledge to "resonate" with oneself or to be attractive from one's perspective?

I argue that we should construe the resonance constraint in terms of affective resonance; something resonates with us when we take pleasure in it. There are a number of reasons why pleasure is a good candidate to explain resonance. Intentional pleasure is a *positive* response to something else, it is an experience that presents other aspects of our lives in a positive way. For instance, we take pleasure *in* or project the positive character of pleasure onto activities like climbing a rock wall and giving to charity. Moreover, when something resonates with us, it should have this sort of positive impression on us where we react to it with an evaluative attitude like pleasure. Pleasure also explains

resonance because when we take pleasure in something, it is difficult to imagine that it might not resonate with us. If I take great pleasure in admiring some artwork or writing a paper, it is hard to see how I might find it uncompelling or unattractive. Enjoying it just is a way in which it is compelling and attractive to me.

Most importantly, I argue that nothing can resonate with us in a way that makes it good for us without our taking pleasure in it. This is because a pleasureless life involves neither resonance thus, *a fortiori*, no well-being. To motivate this idea, consider that a pleasureless life is one in which nothing would seem good to you from within your own subjective perspective; nothing around you would feel like it mattered. This is because pleasure constitutes what it is like for things around you to be experienced as good. So instead, neutral after neutral experience would arise and dissipate without them making the slightest difference to you. You could learn important truths, set and accomplish difficult goals, and interact with family while feeling completely indifferent about these. From your own subjective perspective, these experiences would be a sequence of flavourless perceptions, urges, and intentions that you could care less about. Even if they motivated you, it could not be said that you would care about them in this state, since caring necessarily involves seeing something in a positive light. I find it difficult that such a life could be good for one. I also show that the prudential importance of desire, emotion, and evaluative attitudes rest on their connection to pleasure, without which they seem meaningless. Thus, nothing can resonate or contribute to well-being without our taking pleasure in it. If we accept the affective resonance constraint, this entails a hybrid theory of well-being on which objective states like achievement are good for us when we take pleasure in them.

<div align="right">June 28, 14:15 – 15:00</div>

Eva Rose Read (London School of Economics)
Animal Welfare as Holistic Self-Maintenance: Against the Valence Aggregation Model

An increasingly dominant account of animal welfare is that the only non-instrumental interests of non-human animals are their valenced subjective experiences (VSEs), and that animal welfare is best understood in terms of the aggregate (over some relevant time interval) of their positively and negatively valenced subjective states. I call this view the valence aggregation model of welfare (VAM).

The VAM is the prevailing view in the emerging area of the philosophy of animal welfare (see e.g. Browning 2020, 2023; Browning & Veit 2023), the view held by Effective Altruist groups like Rethink Priorities (see e.g. Fischer 2023) and a view on the rise in animal welfare science (see e.g. Mellor et al. 2020).

The VAM assumes that VSEs matter to other animals. This I grant. The VAM also assumes that VSEs are the only things that matter to other animals. This I argue against.

First, I claim that is in the interests of an animal to have subjective experiences that are accurate reflections of reality when accuracy is part of the function of those subjective experiences. Ignoring this risks perpetuating an impoverished account of what the non-instrumental interests of other

animals are. I argue that this interest does not rely on the individual having a relevant concept or cognitive ability (Browning 2022), but simply on their being sentient.

Second, I argue that it is both the valence and the type of experience that aptly characterises subjective welfare. The VAM, in focussing only on valence, is likely oversimplifying welfare-relevant subjective experience. I find that what is cited as evidence for the commensurability of subjective states according to their valence in welfare, is not such evidence. Drawing on Seth (2021), whose view implies that at least in humans, subjective experiences are aptly distinguished at least in part by their type, or quality, I argue that welfare-relevant subjective experience is better characterised by detail not just about the valence of states, but about which realm of the animal's experience they are relevant to. If the VAM is wrong, in using it we risk wrongfully characterising lives of bad welfare as good by virtue of treating all different subjective states as available for trade-off.

I then propose an understanding of welfare that captures these points, termed welfare as 'holistic self-maintenance'. Here, an animal can do well or badly in their life across multiple domains of self-maintenance. Good welfare is achieved only by a minimal level of flourishing present across all these domains, which itself requires both positive VSEs in these domains, and these being accurate reflections of the animal's extra-subjective reality.

How this alters our current judgements (in using both the VAM and other contemporary definitions of animal welfare) on which animals can reasonably be classed as having 'good welfare' is the focus of future work. But with a broader, enriched view on what matters to other animals, it seems that we will find it increasingly difficult to defend many of their lives under our jurisdiction as 'good'.

References

Browning H. 2020. If I could talk to the animals: measuring subjective animal welfare. PhD Thesis (Australian National University).

Browning H. 2022. The measurability of subjective animal welfare. Journal of Consciousness Studies, 29(3), 150-179.

Browning H. 2023. Welfare comparisons within and across species. Philosophical Studies, 180, 529-551.

Browning H & Veit W. 2023. Positive wild animal welfare. Biology and Philosophy, 38(14), 1-19.

Fischer B. 2023. Rethink Priorities' Welfare Range Estimates. Rethink Priorities. URL: https://rethinkpriorities.org/publications/welfare-range-estimates#:~:text=Given%20hedonism%2C%20an%20individual%27s%20welfare,that%20the%20individual%20can%20realize.

Mellor DJ, Beausoleil NJ, Littlewood KE, McLean AN, McGreevy PD, Jones B & Wilkins C. 2020. The 2020 Five Domains Model: including human-animal interactions in assessments of animal welfare. Animals, 10(10):1870.

Seth AK. 2021. Being you: a new science of consciousness. Penguin.

Theron Pummer (University of St. Andrews)

Hypersensitivity: No Vague Escape

Two years of very intense 'torture' pain is bad. Twenty years of pain 99.9 percent as intense is worse. And two hundred years of pain 99.9 percent as intense as the previous is worse still. We can continue in this fashion. Assuming that at each step there is a sufficient gain in duration and a merely slight drop in intensity, each step seems one for the worse. For some fixed precisification of 'slightly lower intensity', there is a finite number of steps that will lead us to some number of years of very mild 'headache' pain. Since each step is one for the worse, and 'worse than' is transitive, there is some (potentially astronomical) number of years of very mild pain that is worse than two years of very intense pain. This conclusion is counterintuitive. Nonetheless, if pain duration has non-diminishing badness, rejecting any step involves hypersensitivity: a slight decrease in intensity would have to offset any worsening with respect to duration. It has been suggested that appealing to vagueness or indeterminacy allows us to reject a step without hypersensitivity. I show this is false.