Bounded Utilities and Ex Ante Pareto

Petra Kosonen

kosonenpetra@gmail.com

Bounded Expected Totalism

- I will investigate the compatibility of two standard theories: Total Utilitarianism and Expected Utility Theory with a bounded utility function.
- Let's call the combination of these views *Bounded Expected Totalism*.
- I will argue that Bounded Expected Totalism violates *Ex Ante Pareto*, the principle that what is in expectation better for each individual is better overall.

Bounded Expected Totalism

Both Total Utilitarianism and Expected Utility Theory with a bounded utility function are true.

Ex Ante Pareto

For all prospects X and Y, if X is at least as good as Y for each individual, and X is better than Y for some, then X is better than Y.

Total Utilitarianism

- Total Utilitarianism states that a population is better than another just in case the total quantity of well-being it contains is greater.
- The total quantity of well-being in some state of affairs is the sum of individuals' well-being levels in that state of affairs.[1]

Total Utilitarianism

For all states of affairs A and B, A is at least as good as B if and only if the total quantity of well-being in A is at least as great as the total quantity of well-being in B.

[1] Let $\mathrm{W}(A)$ denote the total quantity of well-being in the state of affairs A and let $\mathrm{w}(S_i)$ denote the well-being of individual S_i. Then,

$$\mathrm{W}(A) = \sum_{i=1}^{\infty} w(S_i).$$

Expected Utility Theory

- Expected Utility Theory, in turn, states that a prospect is better than another just in case its expected utility is greater.
- The expected utility of a prospect is calculated by multiplying the utilities of its outcomes by their probabilities, and summing these up.[2]
- Let EU (X) denote the expected utility of prospect X.

Expected Utility Theory

For all prospects X and Y, X is at least as good as Y if and only if EU(X) is at least as great as EU(Y).

[2] Let O be the set of possible outcomes, $p_X(o)$ the probability of outcome o in prospect X and $u(o)$ the utility of o. Then,

$$\text{EU}(X) = \sum_{o \in O} p_X(o)u(o).$$

Boundedness

- What does it mean for utilities to be bounded?

Boundedness

There is some real number M such that for all outcomes x, the utility of x is below M and above $-M$.

- In other words, Boundedness rules out arbitrarily and infinitely good outcomes.
- Standard axiomatizations of expected utility maximization require utilities to be bounded.[3]
- Consider, for example, the von Neumann-Morgenstern axiomatization of Expected Utility Theory.

[3] See for example Kreps (1988, pp. 63–64), Fishburn (1970, pp. 194, 206–207), Hammond (1998, pp. 186–191) and Russell and Isaacs (2021).

Continuity

- Let XpY be a risky prospect such that the agent gets probability p of prospect X and probability $1-p$ of prospect Y
- Then, if prospects are compared by their expected utilities, Boundedness follows from the following von Neumann-Morgenstern axiom:

Continuity

If X is better than Y, which is better than Z, then there are probabilities p and $q \in (0,1)$ such that XpZ is better than Y and Y is better than XqZ.

- For example, suppose a coin is flipped, and an agent gets \$100 with heads and \$0 with tails. Suppose further that it is possible to alter the bias of the coin. Continuity requires that, with some bias, the agent prefers the coin flip to certainly getting \$50 (e.g. 90% chance of \$100 is better than certainty of \$50), but with some other bias, the agent prefers certainly getting \$50 to the coin flip (e.g. 10% chance of \$100 is worse than certainty of \$50).

Continuity implies Boundedness

- In general, Boundedness is false if some prospect A is infinitely better than another (good) prospect B (the two ways in which Boundedness might be false both lead to such prospects given St. Petersburg style cases.)
- This leads to violations of Continuity because the mixed prospect ApC (where C certainly gives nothing) is better than B for all non-zero probabilities p: it is better to get any probability of an infinitely good prospect (and otherwise nothing) than certainty of a finitely good prospect.
- The two ways in which Boundedness might be false (arbitrarily or infinitely good prospects) both lead to violations of Continuity. Therefore, it follows from Continuity that Boundedness is true.[4]
- Next, I will present Bounded Expected Totalism in more detail.

[4] For the proof, see for example Kreps (1988, pp. 63–64).

Bounded Expected Totalism

- Let *well-being* refer to how good some outcome is for an individual.
- And, let *social utility* refer to how good some outcome is overall, from an axiological point of view.
- Also, let *expected individual utility* represent how good some prospect is for an individual, and let *expected social utility* represent how good some prospect is overall.
- In the context of Expected Utility Theory, I will denote these by $\mathrm{EU}_{\mathrm{Ind}}$ and $\mathrm{EU}_{\mathrm{Soc}}$, respectively.
- In general, I will use *individual betterness* to refer to betterness from an individual's point of view.
- Similarly, I will use *overall/impersonal betterness* to refer to betterness from a moral point of view.

The social transformation function

- To combine Total Utilitarianism and Expected Utility Theory, we need a *social transformation function* that takes the total quantity of well-being as input and gives social utilities as output.
- This transformation function must be non-linear if an infinite or arbitrarily large number of happy individuals might exist, as then the total sum of individuals' well-being might be infinite or arbitrarily large (and similarly for negative well-being).
- But, as Bounded Expected Totalism requires expected social utilities to be bounded, the expected social utilities assigned to prospects that might result in an infinite or arbitrarily large number of happy individuals must be bounded.

Finite number of individuals

- One might object that the total quantity of well-being cannot be infinite or arbitrarily large because there is an upper limit to how many individuals might exist, e.g., due to the Universe being finite.
- However, this may not be true, so we need a decision theory that can also handle these possibilities. If there is even a tiny probability that an infinite or arbitrarily large number of individuals exist, then the transformation function must be non-linear for utilities to be bounded.

Probability Fanaticism

- Furthermore, my argument need not rely on infinities.
- One might have a relatively low upper bound for utility to avoid letting outcomes in which very large numbers of happy individuals exist from dominating the expected utility calculations (and similarly for negative well-being).
- Doing so requires a non-linear social transformation function if the total quantity of well-being is very large; the non-linear transformation function makes sure that outcomes containing vast quantities of well-being are assigned relatively low utility numbers.

Probability Fanaticism

- Having an upper bound for utility is one way to avoid *Probability Fanaticism*:[5]

Probability Fanaticism

For any non-zero probability p, and for any (finitely) good outcome o, there is some great enough outcome O such that probability p of O (and otherwise nothing) is better than certainty of o.

[5] Wilkinson (2022, p. 449). For discussions related to fanaticism, see Beckstead (2013, ch. 6), Goodsell (2021), Russell and Isaacs (2021), Wilkinson (2022), Beckstead and Thomas (2023) and Russell (2023).

Non-linear social transformation function

- Suppose that the social transformation function is non-linear. It will also have the following qualities:

1. More well-being is always better, so the social transformation function must be strictly increasing with the total quantity of well-being; it must assign greater utilities to outcomes that contain more well-being.
2. Because utilities are bounded above, similar increases in well-being must (after some point at least) matter less and less. Consequently, the social transformation function must be strictly concave on some subset of its domain. (Similarly: strictly convex if utilities are bounded below).
3. For utilities to be bounded, the social transformation function must be sufficiently concave with positive total well-being and sufficiently convex with negative total well-being; the contribution of additional (positive or negative) well-being to social utility must tend to zero.

Bounded Expected Totalism

- Let f be this transformation function.
- Also, let $p_X(O_i)$ denote the probability of outcome O_i in prospect X and $\mathrm{W}(O_i)$ the total quantity of well-being in O_i.
- Also, recall that $\mathrm{EU}_{\mathrm{Soc}}$ denotes expected social utility.
- Then, we can state Bounded Expected Totalism formally as follows:

Bounded Expected Totalism

Bounded Expected Totalism: For all prospects X and Y, X is at least as good as Y if and only if $\mathrm{EU}_{\mathrm{Soc}}(X)$ is at least as great as $\mathrm{EU}_{\mathrm{Soc}}(Y)$, where

$$\mathrm{EU}_{\mathrm{Soc}}(X) = \sum_{i=1}^{\infty} p_X(O_i) f\left(\mathrm{W}(O_i)\right).$$

Summary

1. On Bounded Expected Totalism, when calculating the value of a prospect, one first calculates the total quantity of well-being in every possible outcome.
2. Then, one transforms each outcome's total quantity of well-being into social utilities.
3. Finally, to get the expected social utility of a prospect, one multiplies the social utility of each outcome with that outcome's probability and sums these up.

- To summarize, social utilities might be bounded if the total quantity of well-being is itself necessarily bounded.
- However, this is not true; therefore, Bounded Expected Totalism requires a social transformation function that takes the total quantity of well-being as input and outputs social utilities.

Bounded Expected Totalism violates Ex Ante Pareto

- Next, I'll give two examples to show that Bounded Expected Totalism violates Ex Ante Pareto if social utilities are bounded above and below.
- These examples show that a violation of Ex Ante Pareto happens regardless of whether individual betterness is risk-neutral, risk-averse or risk-seeking with respect to well-being.
- Consider the following prospects:

Risky vs. Safe:

Risky Gives a 0.5 probability of δ additional well-being for some individual; otherwise, it decreases their well-being by $-\delta$.

Safe Does not increase or decrease well-being.

Risk-aversion

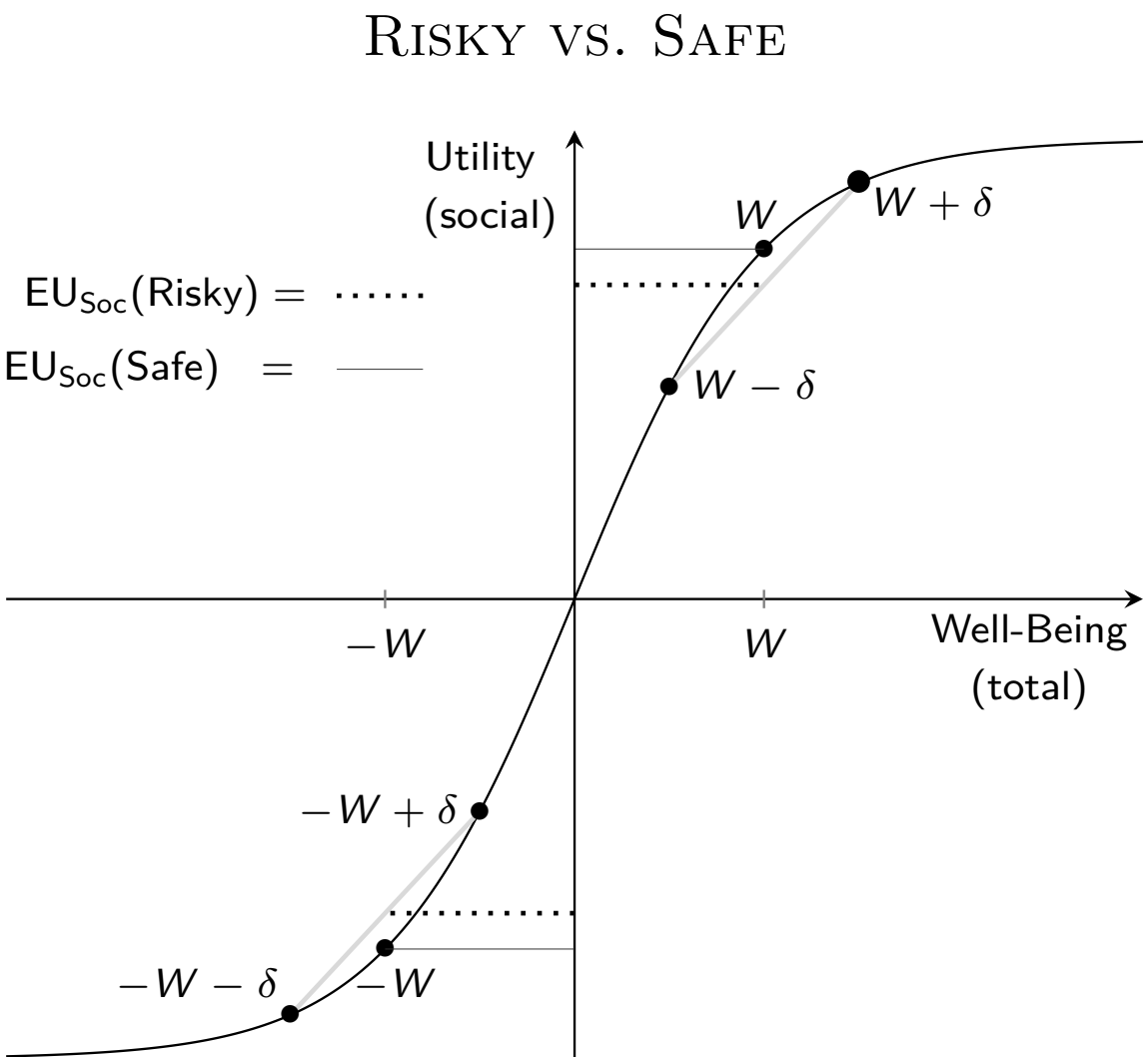
- If social utilities are bounded above, then (at least at some point) the social transformation function is concave with a positive total quantity of well-being.
- This means that, at least sometimes, the overall betterness relation is risk-averse with respect to well-being.
- So, with some positive total quantity of well-being W, Safe is impersonally better than Risky.

Risk-seeking

- On the other hand, if social utilities are bounded below, then (at least at some point) the social transformation function is convex with a negative total quantity of well-being.
- This means that, at least sometimes, the overall betterness relation is risk-seeking with respect to well-being.
- Thus, with some negative total quantity of well-being $-W$, Risky is impersonally better than Safe.

Graph

- This is illustrated by the following graph:



Risky vs. Safe

- So, whether Risky is overall better than Safe (or vice versa) depends on the total quantity of well-being.
- However, whether Risky is better than (or equally as good or worse than) Safe for some agent Alice does not depend on the total quantity of well-being.

Individual betterness is risk-seeking

- First, suppose that Risky is better than Safe for Alice (and no one else is affected by the choice).
- Then, Bounded Expected Totalism violates Ex Ante Pareto when the total quantity of well-being in the background population is high (W) because then Safe is better than Risky impersonally.
- In that case, individual betterness for Alice is risk-seeking with respect to well-being, but overall betterness is risk-averse.

Individual betterness is risk-averse

- Next, suppose that Safe is better than Risky for Alice (and again equally good for everyone else).
- Then, Bounded Expected Totalism violates Ex Ante Pareto when the total quantity of well-being in the background population is low ($-W$) because then Risky is better than Safe impersonally.
- In that case, individual betterness for Alice is risk-averse, but overall betterness is risk-seeking.
- So, this case shows that Bounded Expected Totalism violates Ex Ante Pareto if individual betterness for Alice deviates from risk-neutrality.

Weak Ex Ante Pareto

- The previous case also shows that Bounded Expected Totalism violates Weak Ex Ante Pareto if individual betterness is risk neutral.
- Suppose that Risky and Safe are equally good for Alice.
- This means that individual betterness for Alice is risk-neutral with respect to well-being.
- This time we get a violation of Weak Ex Ante Pareto:

Weak Ex Ante Pareto

For all prospects X and Y, if X is at least as good as Y for each individual, then X is at least as good as Y.

Weak Ex Ante Pareto

- Given that Risky and Safe are equally good for Alice, Risky is at least as good as Safe and Safe is at least as good as Risky for Alice.
- However, when the total quantity of well-being in the background population is W, Risky is not at least as good as Safe impersonally (because then Safe is better than Risky impersonally).
- Similarly, when the total quantity of well-being in the background population is $-W$, Safe is not at least as good as Risky impersonally (because then Risky is better than Safe impersonally).
- So, the previous case shows that Bounded Expected Totalism violates *Weak Ex Ante Pareto* if individual betterness for Alice is risk-neutral.

The Risk-Neutral Case

- We may change the previous case slightly to show that a violation of Ex Ante Pareto happens again:[6]

  The Risk-Neutral Case:

  Risky Gives a 0.5 probability of $\delta + \epsilon$ additional well-being for some individual; otherwise, it decreases their well-being by $-\delta$.

  Safe Does not increase or decrease well-being.

[6] Gustafsson (2022) presents a similar case to illustrate that *Ex-Post* Prioritarianism violates Ex Ante Pareto, a fact that goes back at least to Rabinowicz (2002). For an overview of this topic, see for example Fleurbaey (2018). See also Broome (1991, Ch. 9). Bounded Expected Totalism applies the transformation function to the total quantity of well-being; *Ex-Post* Prioritarianism applies it to the well-being of individuals.

The Risk-Neutral Case

- Given that individual betterness for Alice is risk-neutral, Risky is better than Safe for Alice (with all positive values of δ and ϵ).
- However, given that overall betterness is risk-averse when the total quantity of well-being in the background population is high (W), we can find some positive constants δ and ϵ such that Safe is impersonally better than Risky when the total quantity of well-being in the background population is W.
- Then, Risky is better than Safe for Alice (and equally good for everybody else) but Safe is better than Risky impersonally.
- So, we have a violation of Ex Ante Pareto.

More formally

- More formally, given that the social transformation is concave at W, there must be some positive constants δ and ϵ such that $f(W) - f(W - \delta) > f(W + \delta + \epsilon) - f(W)$.
- This is because the smaller benefit (δ) contributes more when added to a population at a lower well-being level than the greater benefit ($\delta + \epsilon$) when added to a population at a higher well-being level.
- The expected social utility of Safe is $\text{EU}_{\text{Soc}}(\text{Safe}) = f(W)$.
- And, the expected social utility of Risky is $\text{EU}_{\text{Soc}}(\text{Risky}) = \frac{1}{2} \cdot f(W + \delta + \epsilon) + \frac{1}{2} \cdot f(W - \delta)$.
- Given that $f(W) - f(W - \delta) > f(W + \delta + \epsilon) - f(W)$, $\text{EU}_{\text{Soc}}(\text{Risky})$ is less than $\text{EU}_{\text{Soc}}(\text{Safe})$.[7]

[7] By rearranging $f(W) - f(W - \delta) > f(W + \delta + \epsilon) - f(W)$, we get $f(W) + f(W) > f(W + \delta + \epsilon) + f(W - \delta)$. Next, by dividing both sides by 2, we get $f(W) > \frac{1}{2} \cdot f(W + \delta + \epsilon) + \frac{1}{2} \cdot f(W - \delta)$.

Summary

- To summarize, I presented two examples to show that Bounded Expected Totalism violates Ex Ante Pareto regardless of whether individual betterness is risk-neutral, risk-averse or risk-seeking with respect to well-being.[8]

[8] The latter case shows that Bounded Expected Totalism violates Ex Ante Pareto if individual betterness is risk-neutral and utilities are bounded above. One can change the case to show that Ex Ante Pareto is also violated if individual betterness is risk-neutral and utilities are bounded below.

Harsanyi's social aggregation theorem

- Next, I will discuss how these examples relate to a famous result in this area, namely, Harsanyi's social aggregation theorem.
- Harsanyi's social aggregation theorem shows that if both individual and social betterness relations can be given an expected utility representation, and the overall betterness relation satisfies Ex Ante Pareto, then social utilities are weighted sums of individual utilities.[9]
- Premises:

1. Each individual's betterness relation obeys the von Neumann-Morgenstern axioms.
2. The overall betterness relation obeys the von Neumann-Morgenstern axioms.
3. Ex Ante Pareto.[10]

[9] Harsanyi (1955).

[10] Harsanyi (1955) uses Pareto Indifference in the original formulation of the theorem, while Harsanyi (1977, p. 65) uses Weak Ex Ante Pareto.

Harsanyi's social aggregation theorem

- The conclusion of Harsanyi's theorem is that social utilities are weighted sums of individual utilities.
- Thus, overall betterness can be represented as maximizing the expectation of a weighted sum of individual utilities.
- If, in addition, we assume equal weighting for all individuals, then this theorem shows that the social utility function must be a sum of individual utilities.
- Harsanyi's theorem shows, in other words, that if individual and overall betterness relations are represented by expectational utility functions, then in order to satisfy Ex Ante Pareto, the social utility function must be a linear combination of individual utilities.

Harsanyi's social aggregation theorem

- I showed that Total Utilitarianism combined with bounded Expected Utility Theory violates Ex Ante Pareto.
- Therefore, if one accepts Bounded Expected Totalism, that premise of Harsanyi's theorem fails.
- The reason that led to its failure was that a non-linear social transformation function is needed 1.) because the number of individuals might be infinite or arbitrarily large and 2.) to avoid letting outcomes in which very large numbers of happy individuals exist from dominating the expected utility calculations.

Harsanyi's social aggregation theorem

- In fact, it is unsurprising that one of Harsanyi's premises must be rejected; if the number of individuals might be infinite or arbitrarily large, then social utilities cannot be weighted sums of individual utilities because this could lead to unbounded social utilities.
- Similarly, if one uses a non-linear transformation function to avoid Probability Fanaticism, then social utilities cannot be weighted sums of individual utilities because these sums might exceed the upper bound of utilities.
- So, given that a bounded expected totalist rejects Harsanyi's conclusion, they cannot accept all his premises.

Harsanyi's social aggregation theorem

- This is worrying because Harsanyi's theorem is often considered one of the best arguments for utilitarianism.
- The conclusion of Harsanyi's theorem is that, for any fixed and finite population, social utility is an affine (or linear) function of total individual utility.
- However, once we consider the possibility of an infinite or arbitrarily large population (or wish to avoid fanatical prescriptions), we find that social utility must be non-linear if social utilities are bounded and additive with individual utilities.
- And this leads to violations of Ex Ante Pareto.

Harsanyi's social aggregation theorem

- These cases might be taken to undermine Boundedness (and Continuity). One might accept, for example, *Unbounded Expected Totalism*, namely, the view that combines Total Utilitarianism and Expected Utility Theory with an unbounded utility function.[11]
- Alternatively, the arguments might be taken to indirectly support alternatives to Boundedness, such as discounting small probabilities (also violates Ex Ante Pareto).
- Given that both Bounded Expected Totalism and discounting small probabilities violate Ex Ante Pareto, the plausibility of Ex Ante Pareto does not favor the former over the latter.

[11] See for example McCarthy et al. (2020).

Conclusion

- I have shown that Bounded Expected Totalism violates Ex Ante Pareto.
- I gave separate examples of Ex Ante Pareto violations for risk-neutrality and risk-aversion/risk-seeking.
- I also discussed the implications of this case for Harsanyi's social aggregation theorem.
- To conclude, combining two standard theories, Total Utilitarianism and Expected Utility Theory with a bounded utility function, results in violations of Ex Ante Pareto: The combination of these views implies that a prospect can be impersonally better than another prospect even though it is worse for everyone who is affected by the choice.[12]

[12] This talk focused on the compatibility of Expected Utility Theory and Total Utilitarianism, but the problem with Ex Ante Pareto arises for, for example, Critical-Level Utilitarianism in exactly the same way.

References I

Beckstead, N. (2013), On the overwhelming importance of shaping the far future, PhD thesis, Rutgers, the State University of New Jersey.

Beckstead, N. and Thomas, T. (2023), 'A paradox for tiny probabilities and enormous values'.

Broome, J. (1991), *Weighing Goods: Equality, Uncertainty and Time*, Blackwell, Oxford.

Fishburn, P. C. (1970), *Utility Theory for Decision Making*, Wiley, New York.

Fleurbaey, M. (2018), 'Welfare economics, risk and uncertainty', *Canadian Journal of Economics* **51**(1), 5–40.

Goodsell, Z. (2021), 'A St Petersburg Paradox for risky welfare aggregation', *Analysis* **81**(3), 420–426.

Gustafsson, J. E. (2022), 'Ex-ante prioritarianism violates sequential ex-ante Pareto', *Utilitas* **34**(2), 167–177.

References II

Hammond, P. J. (1998), Objective expected utility: A consequentialist perspective, *in* S. Barberà, P. J. Hammond and C. Seidl, eds, 'Handbook of Utility Theory Volume 1: Principles', Kluwer, Dordrecht, pp. 143–211.

Harsanyi, J. C. (1955), 'Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility', *Journal of Political Economy* **63**(4), 309–321.

Harsanyi, J. C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, Cambridge.

Kreps, D. M. (1988), *Notes on the Theory of Choice*, Westview Press, Boulder.

McCarthy, D., Mikkola, K. and Thomas, T. (2020), 'Utilitarianism with and without expected utility', *Journal of Mathematical Economics* **87**, 77–113.

Rabinowicz, W. (2002), 'Prioritarianism for prospects', *Utilitas* **14**(1), 2–21.

References III

Russell, J. S. (2023), 'On two arguments for fanaticism'.

Russell, J. S. and Isaacs, Y. (2021), 'Infinite prospects', *Philosophy and Phenomenological Research* **103**(1), 178–198.

Wilkinson, H. (2022), 'In defence of fanaticism', *Ethics* **132**(2), 445–477.