

# Value Theory Seminar: HT14 Week 1

## Core reading (mandatory)

Those attending the first class should pre-read the following papers for discussion:

- Parfit, D. (1997). Equality and priority. *Ratio*, 10(3), 202-221.
- Otsuka, M., & Voorhoeve, A. (2009). Why it matters that some are worse off than others: an argument against the priority view. *Philosophy & Public Affairs*, 37(2), 171-199.
- Greaves, H. (MS) Antiprioritarianism.

All of these papers are available on Weblearn (navigate to 'Seminar Resources' → 'Hilary Term 2014' → 'Value Theory' → 'Week 1').

## Further reading (optional)

Those more familiar with the debates and/or with more time might also like to look at some of the following:

Nagel, T. (1979). Equality. In his *Mortal questions*, Cambridge University Press.

- *An even earlier statement sharing many intuitions with Parfit's 'Equality and priority', but with important points of difference.*

Broome, *Weighing Goods*, OUP 1991.

- *Critical of prioritarianism: follow up the references to 'prioritarianism' in the index.*

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

- *Suggests a way for prioritarians to respond to Broome's criticisms.*

Any or all of the papers in a special issue of *Utilitas* devoted to Otsuka and Voorhoeve's paper: Volume 24, Special issue 03 (Prioritarianism), September 2012.

- *Many of these (the articles by Porter, Parfit, Williams and O'Neill) are sympathetic to prioritarianism, and focus on defending it against Otsuka and Voorhoeve's criticism.*

Greene, J., & Baron, J. (2001). Intuitions about declining marginal utility. *Journal of behavioural decision making*, 14 , 243-55.

- *Reports a psychological experiment that arguably casts doubt on 'prioritarian' intuitions.*

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

- *Original article setting out an important and much-neglected theorem that is sometimes claimed to be a 'proof' of utilitarianism (as against e.g. prioritarianism). (For an alternative exposition, see Broome, *Weighing Goods*, chapter on 'the interpersonal addition theorem'. An extremely brief description of the result itself – not its proof – is contained in Greaves' 'Antiprioritarianism'.)*



**Utilitarianism and Prioritarianism**

1. A perennial complaint against utilitarianism: utilitarianism cares only about the *total amount* of welfare, but (intuition!) the *distribution* of welfare among people is important too.
  - a. Nagel's example: Moving to the city or the suburb
2. Two types of distribution-sensitive view: egalitarianism and prioritarianism
  - a. (Telic, Welfare-)Egalitarianism: Inequality between the welfare levels of different people is intrinsically bad. ('Intrinsically' as opposed to merely instrumentally: cf. divided-world scenarios.)
  - b. Prioritarianism: "Benefitting people matters more the worse off they are." That is: a *benefit of a given size* improves the overall state of affairs more if it is given to a worse off person than if it is given to a better off person.
  - c. Teasing apart prioritarianism and egalitarianism
    - i. Levelling down?
    - ii. Parfit's analogy to breathing at altitude.
  - d. The value-function approach
    - i. The approach: write down a value function that takes individuals' well-being levels as inputs and outputs a number representing 'overall goodness': the higher that number, the better the state of affairs in question.
    - ii. The P/E distinction, in this approach: Prioritarianism postulates a value function exhibiting *additive separability of persons*; egalitarianism has a non-separable value function.
3. The 'content problem': To make the prioritarian claim precise, the notion of 'benefit of a given size' is crucial. But what is it for one benefit to one person to be the 'same size' as another benefit to another person?
  - a. 'Increase his well-being by the same amount' is not an answer: just a translation of the problem. The question is how to 'cardinalise' the well-being scale.
  - b. Possibility 1: Benefit 1 to Person 1 improves the overall state of the world by the same amount that Benefit 2 to Person 2 does. (Renders prioritarianism trivially false.)
  - c. Possibility 2: Just a primitive fact of the matter! (Illegitimate postulate, even if one is not a positivist.)
  - d. Possibility 3: Appeal to decision theory for the cardinalisation...
4. Decision-theoretic approach: Suppose that in addition to a ranking of states of affairs in terms of how good they are for a given person, we also have a ranking of lotteries (probability distributions over states of affairs) in terms of how good they are 'ex ante' for a given person. The representation theorems of decision theory then show that provided such a ranking (of lotteries) obeys a number of intuitively reasonable principles, the ranking has an expected utility representation, with a unique probability distribution, and a utility function that is *unique up to positive affine transformation*. The latter feature (ignoring the problem of interpersonal comparisons, for now!) gives us a 'cardinal' notion of well-being.
5. Otsuka and Voorhoeve's objection to ("Technical") prioritarianism: Prioritarianism places the "moral shift" in the wrong place.
  - a. O&V: analogous intrapersonal (inter-state of nature, i.e. uncertainty) and interpersonal (intra-state of nature) cases should be treated differently.

- i. Example: (1) 'slight impairment' → 'perfect health' vs (2) 'very severe impairment' → 'severe impairment'
  - ii. Q1: You have a 50:50 chance of suffering from each condition (1), (2). Which would you rather receive treatment for?
  - iii. Q2: You are a morally motivated stranger, answering Q1 on behalf of a third party.
  - iv. Q3: You are a morally motivated stranger, deciding the analogous question for a *group* of people, half of whom suffer from each condition.
- b. O&V hold that if the answer to Q1 is "indifferent", then *the answer to Q2 is indifference and the answer to Q3 is "treat the very severe impairment"*. But prioritarianism doesn't give us this: it places the "shift" (instead) between Q1 and Q2.
  - i. And placing the "shift" here amounts to violating the "ex ante Pareto" principle (even) in one-person cases. This is crazy.
- c. What's the right theory, then?
  - i. Egalitarianism? But then we're back to the Levelling Down objection.
  - ii. A contractualist approach? "[I]n any scenario involving more than one person, the allocation of resources must be justifiable to each person taken separately in a manner that brings interpersonal considerations to bear that cannot apply in the case of one person considered in isolation." (?? → Wk 5)
- d. Prioritarian responses to Otsuka and Voorhoeve
  - i. Deny the intuition that supports the special case of Ex Ante Pareto? Crazy.
  - ii. Limit the scope of the prioritarian theory? Ad hoc.
  - iii. Autonomy? Irrelevant if the patient has no *right* to the treatment anyway.
  - iv. Needs and personal projects? This doesn't cover enough cases, and anyway wouldn't deliver the "shift" that O&V insist on.
  - v. Go deontic? Let's settle the axiological question first.
- 6. Greaves' objection to Technical Prioritarianism
  - a. *Technical* prioritarianism is (anyway) just nuts: by reframing the questions, one can just as well argue for "Technical anti-prioritarianism". Something has gone wrong in prioritarian theory long before O&V came along.
- 7. A theorem worth knowing: Harsanyi's aggregation theorem
  - a. The theorem: suppose that
    - i. The individual ex ante lottery rankings respect decision theory.
    - ii. The overall ex ante lottery ranking respects decision theory.
    - iii. The 'Ex ante Pareto principle' is satisfied.
 Then the overall value function is a (weighted) sum of the individuals' von Neumann-Morgenstern utility functions (i.e., "Technical utilitarianism" is true).
- 8. An experiment worth knowing: Greene and Baron's "Intuitions about declining marginal utility"
  - a. Subjects report intuitions of declining marginal utility not only of money, years of life etc, but even *of utility*. (Fodder for a debunking argument against the prioritarian (and perhaps also the welfare-egalitarian?) intuition.)

## **Value Theory Seminar: HT14 Week 3**

Readings marked “\*” are available on Weblearn.

### **Core reading (mandatory)**

Robbins, *On the nature and significance of economic science*, pp.120-5 in the first edition (MacMillan, 1932).

*Classic statement of scepticism w.r.t. the “scientific” status of interpersonal utility comparisons (IPUCs).*

Harsanyi, *Rational behaviour and bargaining equilibrium in games and social situations*, Cambridge university press, 1977; sections 4.1-4.4 (pp.48-60).

*Sets out Harsanyi’s attempt to ground IPUCs via the notion of “extended preferences”.*

\*Broome, ‘Extended preferences’, in C. Fehige and U. Wessels, eds., *Preferences*, pp. 271–287. Berlin: Walter de Gruyter.

*Criticism of Harsanyi.*

\*Adler, Well-being and fair distribution: Beyond cost-benefit analysis, chapter 3, sections II (“Interpersonal comparisons”), III (“Well-being: an account”).

*An attempt to rehabilitate the extended-preferences approach in the light of several criticisms of Harsanyi’s particular way of implementing it (including Broome’s).*

If you have time to add one more short thing to your reading list, a good next step is the extract from Broome’s “Weighing lives” cited below.

### **Further reading (optional)**

#### **Further classic (and short!) statements on IPUCs in the history of economics:**

- Jevons, *The theory of political economy*, pp.13-16 in the second edition, Macmillan 1879. Available via <https://www.dropbox.com/s/vw0r5shyw7degli/jevons%20-%20theory%20of%20political%20economy.pdf>
- \*Robbins, ‘Interpersonal utility comparisons: a comment’, *The Economic Journal*, Vol. 48, No. 192 (Dec., 1938), pp. 635-641.
- \*Harrod, pp. 395-7 in his ‘Scope and method of economics’, *The Economic Journal*, Vol. 48, No. 191 (Sep., 1938), pp. 383-412.
- \*Kaldor, ‘Welfare propositions of economics and interpersonal comparisons of utility’, *The Economic Journal*, Vol. 49, No. 195 (Sep., 1939), pp. 549-552.

**More technical material on the ‘Paretian’ approach to cost-benefit analysis**, which attempts to do without IPUCs (building on Kaldor 1939): this is useful background (if you can follow it) by way of *motivation* for making IPUCs – “look how bad things get if you don’t have IPUCs” – but has nothing to say about our central question of *how* to make/give content to them. The final piece (Ord) is a short non-technical summary of the issues.

- Hicks, J. R., 'The foundations of welfare economics', *Economic Journal*, 49 (1939), pp. 697–712.
- de Scitovszky, Tibor, 'A note on welfare propositions in economics', *Review of Economic Studies*, 9 (1941), pp. 77–88.
- Gorman, W. M., 'The intransitivity of certain criteria used in welfare economics', *Oxford Economic Papers*, 7 (1955), pp. 25–35.
- Boadway R. (1982). The welfare foundations of cost-benefit analysis. *Economic Journal* 84, 926–939.
- \*Blackorby C., and D. Donaldson (1990). A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *Canadian Journal of Economics*, 471–494.
- \*Ord, 'Problems with Hypothetical Pareto Improvements'

**Misc:**

- Broome, *Weighing Goods*, pp.219-20
- \*Broome, *Weighing Lives*, section 5.3 ("Comparing well-being between people", pp.91-7).
- Scanlon, 'The moral basis of interpersonal comparisons', in Elster and Roemer (eds.), *Interpersonal comparisons of well-being*, CUP 1991.
- Arrow, section V of his 'Extended sympathy and the possibility of social choice', *American economic review* 1977, pp 219-225. (*More on the extended-preference approach: Arrow argues in defence of Harsanyi's claim that all individuals' extended-preferences will agree with one another.*)

**Interpersonal comparisons of utility (IPUCs)**

1. Skepticism about IPUCs
  - a. Observations of individual behaviour plausibly suffice to determine individuals' *separate* utility functions up to positive affine transformation. But nothing *in individual behaviour, or any other observable fact*, seems to fix interpersonal comparisons of utility.
    - i. Interpersonal *level* comparisons: Is person *i* better off than person *j*?
    - ii. Interpersonal *unit* comparisons: Does the change from state of affairs A to state of affairs B make *more difference* to person *i* or to person *j*? (Or: What is the *ratio* of the difference it makes to A to the difference it makes to B?)
  - b. Examples: Daily washing up, Robbins' Brahmin
  - c. The positivist conclusion from this: IPUCs are *meaningless*.
  - d. A more moderate (and plausible) account of meaningfulness: the meaning of a ("theoretical") term is fixed by the theory-network in which that term is embedded, together with the connections (input-causal, predictive, evaluative) between that network and the world.
  - e. More moderate conclusion from the above observations: the content of an IPUC-claim is *evaluative*, not descriptive. (Robbins: and therefore IPUCs have no place in *positive* economics.) E.g. as in
  - f. **Account 1: IPUCs from overall betterness:** to say that  $A \rightarrow B$  increases *i*'s well-being more than it decreases *j*'s well-being (just) is to say that, other things equal, B is better-overall than A.
    - i. Objection 1: this gets explanation backwards
    - ii. Objection 2: Even if *true*, this statement isn't *analytic* (Broome, WL)
2. Normative economics without IPUCs: The "Paretian" approach
  - a. The "Paretian approach"/the "new welfare economics" attempts to do even *normative* economics without IPUCs. (Because of positivism? Or epistemological worries?)
  - b. The Pareto criterion:  $B > A$  if for all *i*,  $B >_i A$ .
    - i. A.k.a. the "person-affecting claim/principle". Arguably (i.e. modulo issues of desert, etc) a *sufficient* condition for  $B > A$ , but certainly not *necessary*.
  - c. The 'Potential Pareto' criterion:  $B > A$  if the sum of "compensating variations" involved in the transition  $A \rightarrow B$  is positive.
    - i. The "compensating variation" for person *i* associated with the transition  $A \rightarrow B$ : How much would person *i* be willing to pay to have this transition go ahead? (In terms of utility:  $u_i(A, m_i^A) = u_i(B, m_i^A - CV_{i, A \rightarrow B})$ .)
    - ii. The idea behind the PPC is: if the PPC is satisfied, then if the transition  $A \rightarrow B$  goes ahead, the 'winners' from this transition *could* compensate the losers in such a way that the combined transition satisfies the original Pareto criterion. The "pie" has been "grown".
  - d. Ethical criticism of the PPC
    - i. This amounts to giving the interests of the rich much more weight than those of the poor. (Applications to climate change.)

- ii. It's crazy to think that the transition is an improvement *even if compensation is not paid*. Example: a punch on the nose.

### 3. Account 2: IPUCs via "extended preferences"

- a. Harsanyi's version: Assume that each individual has preferences over *extended alternatives*: pairs consisting of (A, P) a centred possible world A and a set P of personal preferences.
- b. Criticism of Harsanyi
  - i. It's just *not true* that every individual will have the same extended preferences. Harsanyi's argument to the contrary is confused. (Broome, EP)
  - ii. Metaphysical scruples: Harsanyi's interpretation of the extended-preference orderings is in terms of "whether [individual i] would prefer to be [as he is or] in objective position  $B_j$  with j's subjective attitudes  $P_j$ " (p.52). But it is metaphysically impossible for i to be in objective position  $B_j$ . So this is incoherent. (Adler)
  - iii. Harsanyi uses relatively unlauded preferences. But it is a familiar point from discussions of the preference-satisfaction theory of well-being that (for evaluative/normative purposes) *irrational* preferences must be excluded. (Adler)
- c. Adler's version: Take the "preferences" in question to be individuals' *judgments* as to whether i is better off in A than j is in B (thus dealing with (ii)). Require the preferences to be (procedurally) rational (thus dealing with (iii)). Accept that individuals will still have different extended-"preference" orderings: recover betterness-overall by supervalueing.
- d. Criticism of Adler:
  - i. The supervalueational approach leads to massive (indeed, possibly total) incomparability.
  - ii. Once we've replaced preferences with judgments of betterness-for-individuals, why not just theorise directly about betterness-for-individuals itself?

### 4. Account 3: IPUCs via betterness-for-the-individual

- a. Version 1 of this account: Write  $A \succ_i B$  for "A is better-for-i than B" (A, B possible worlds). Extend  $\succ_i$  to *lotteries* over possible worlds. Require  $\succ_i$  to obey EU theory. Require in addition that  $\succ_i, \succ_j$  are related as follows: if possible world  $\pi_{ij}(A)$  is obtained from A by permuting individuals i and j, then  $A \succ_i B$  iff  $\pi_{ij}(A) \succ_j \pi_{ij}(B)$ . It follows that each  $\succ_i$  is (as usual) expectationally represented by a cardinal family of utility functions, *and* that i's family of utility functions is obtained from j's by composition with  $\pi_{ij}$ . This last means that there is a privileged way to make IPUCs. (Basic idea here: once  $\succ_i$  is taken to represent betterness-for-i rather than i's preferences, it is overwhelmingly plausible that the orderings for different individuals are "identical", in which case IPUCs are unproblematic.)
- b. A worry: Version 1 requires haecceitism. (Cf. Adler's criticism (ii) of Harsanyi.)
- c. Version 2 (non-haecceitistic): Admit a *single* betterness-for-the-individual ordering,  $\succ_{\text{indiv}}$ .  $\succ_{\text{indiv}}$  orders *centred* worlds: write  $(A,i) \succ_{\text{indiv}} (B,j)$  for "being i in world A is better-for-the-individual than being j in world B". Extend  $\succ_{\text{indiv}}$  to lotteries on centred

worlds<sup>1</sup>, and require it to obey EU theory. Then there exists a cardinal family of utility functions  $U_{\text{indiv}}$  that expectationally represent  $>_{\text{indiv}}$ ; and the restriction  $U_i$  of  $U_{\text{indiv}}$  to worlds centred on  $i$  is “ $i$ ’s utility function” in the usual sense. Any  $U_{\text{indiv}}$  in this cardinal family trivially defines IPUCs. (The basic idea here is to enlarge the informational base, so that – unlike the collection of individuals’ separate orderings  $>_i$  – the ‘master ordering’  $>_{\text{indiv}}$  from which we start already ‘tells’ us the IPUC facts.)

- d. These accounts do make IPUCs evaluative in content.
- e. But they do not lead to the charges of explanatory reversal and/or false analyticity that we levelled at Account 1.

#### 5. Account-type 4: IPUCs via a particular substantive theory of well-being

- a. Commitment to a particular theory of well-being might give us additional resources for saying more things about IPUCs. Will their content then still be ‘evaluative’? Some thoughts...
- b. Hedonism and objective-list theory: Ironically, seem to render *intrapersonal* comparisons ‘evaluative’, but *interpersonal* ones ‘descriptive’.
- c. Preference-satisfaction theory: Intrapersonal comparisons have a strong ‘descriptive’ component, but also an ‘evaluative’ component insofar as preferences must be laundered for (substantive) rationality. Interpersonal comparisons are ‘descriptive’ insofar as there is an empirical matter of fact concerning relative interpersonal strength of preferences.
- d. Nothing that any substantive account of well-being says about IPUCs is likely to be *analytic*. (Does this matter?)

#### 6. Conclusions

- a. IPUCs are not *meaningless*. (Only an implausible positivism implies otherwise.)
- b. Their content is certainly *partly* evaluative, and may be wholly so.
- c. But they are not alone in that (*intrapersonal comparisons of the sort of ‘utility’ that is relevant for normative purposes* are evaluative in content too).
- d. And their evaluative status is no reason not to use them in normative economics.
  - i. The Paretian approach is (in light of its immense drawbacks) insufficiently motivated.

---

<sup>1</sup> We can exclude “mixed” lotteries here, without blocking the theorem we need, and might want to for interpretational reasons.

## **Value Theory Seminar: HT14 Week 5**

*Readings marked “\*” are available on Weblearn.*

### **Core reading (mandatory)**

Scanlon, *What we owe to each other*, ch 5, sec. 9 (“Aggregation”).

\*Kamm, ‘Aggregation and two moral methods’

\*Otsuka, ‘Saving lives, moral theory, and the claims of individuals’

### **Further reading (optional)**

\*Carlson, ‘Should we kill to avoid headaches?’

Taurek, ‘Should the numbers count?’

Parfit, ‘Innumerate ethics’

Anscombe, ‘Who is wronged?’

\*Munoz-Darde, ‘The distribution of numbers and the comprehensiveness of reasons’

Timmerman, ‘The individualist lottery: how people count, but not their numbers’

\*Otsuka, ‘Skepticism about saving the greater number’

\*Otsuka, ‘Scanlon and the claims of the many versus the one’

\*Broome, ‘Fairness’

**Complaint-aggregation; the “separateness of persons” objection to utilitarianism**

1. The discontents of utilitarianism, part II: The “separateness of persons”
  - a. Problem(?): Utilitarianism recommends gladiator-style entertainments in which the trivial enjoyments of many outweigh extreme harms to a few, provided that the “many” are numerous enough.
  - b. Claim-theorists seek a different fix – one that they take to be independently motivated:
    - i. “[Contractualism insists] that the justifiability of a moral principle depends only on various *individuals*’ reasons for objecting to that principle and alternatives to it... [This] blocks [utilitarianism’s reasoning regarding gladiator-style cases] in an intuitively appealing way. It allows the intuitively compelling complaints of those who are severely burdened to be heard, while, on the other side, the sum of the smaller benefits to others has no justificatory weight, since there is no individual who enjoys these benefits and would have to forgo them if the policy were disallowed.” (Scanlon, *WWOTEO*, p.230; emphasis added)
    - c. Our question: Is there really any viable theory to be built on this “individualist restriction”/this talk of separateness of persons/of claims and complaints? Or is the claim-theorists’ rhetoric in the end empty?
2. The foil: Two varieties of extreme non-aggregationism
  - a. Extreme non-aggregationism in Taurekian “save the greater number?” scenarios
    - i. Suppose that you are out at sea in your pleasure yacht. There is one stranger stranded on Rock A, and there are five strangers stranded on Rock B, all threatened by the rising tide. You have time to rescue *either* the one on rock A *or* the five on Rock B, but not both.
    - ii. A spectrum of views on “numbers” cases
      1. Permissive view: You are morally permitted to save the One, and similarly the Five. (Anscombe)
      2. Simple lottery view: You are obliged to flip a fair coin. (Taurek)
      3. Weighted lottery view: You are obliged to roll a fair die and save the One iff the die lands 6. (Timmerman)
      4. Utilitarian’s view: You are morally required to save the Five.
    - iii. Claim-theorists typically (want to) *agree with the utilitarian* on these “numbers” cases.
  - b. A strong individualist restriction *on the overall-betterness ordering*
    - i. **Method 0:** If we refuse to make *any* trade-offs between benefits accruing to some parties and harms accruing to others (the “strong interpretation of the separateness of persons objection to utilitarianism” – cf Brink, “The separateness of persons, distributive norms, and moral theory”), then our overall ranking of states of affairs will exhibit massive incomparability. We are left only with the Pareto/person-affecting principle. (Cf. wk. 3.)
    - ii. **Method 0.5:** The “anonymous Pareto” principle discussed by Otsuka is (a bit, but) little better.
3. A contractualist approach?

- a. The idea of a “complaint”: In the context of a choice between two policies or courses of action, A and B say, the ‘size’ of a given person’s complaint against policy A might be: the *amount by which her well-being under A falls short of* her well-being under B.
    - i. NB Usually not the contractualists’ full notion of “complaint”, but should(?) suffice for our current purposes.
  - b. The *strength* of a complaint:
    - i. Size-only version: the “strength” of a complaint just is its size.
    - ii. Complaint-size plus baseline version: Define the strength of a complaint to be the size of that complaint, multiplied by some factor that is a decreasing function of the absolute well-being level that the person has in the scenario in which she is worse off.
    - iii. Baseline-only version: Define the strength of a complaint to be some decreasing function of the *absolute well-being level* that the person has in the scenario in which she is worse off.
    - iv. (Cf. the distinction between utilitarianism, prioritarianism and maximin)
  - c. **Method 1:** One ought to pursue whichever policy has the strongest *individual* complaint in its favour. (Kamm: the “blinker pairwise comparison” method.)
  - d. Worry: Regardless of which notion of complaint strength is used, *this* approach cannot ground the conclusion that one ought to save the Five rather than the One in Taurek-style “numbers” cases.
4. “Balancing”/“tiebreaker” aggregation
- a. **Method 2:** Second “pairwise-comparison” method (Kamm’s terminology: the “context-aware pairwise comparison” method)
    - i. When comparing policies 1 and 2, first divide all affected individuals into two groups: Let  $S_1$  (resp.  $S_2$ ) contain all and only those who would fare better under 1 (resp. 2).
    - ii. Suppose there exists a one-to-one function  $f$  from  $S_1$  into  $S_2$ , such that:
      - 1. For all  $i$  in  $S_1$ , the complaint that person  $i$  has if policy 2 is followed is of equal or lesser strength than the complaint that person  $f(i)$  has if 1 is followed.
      - 2. In addition, either there is some person  $i$  such that the complaint that  $i$  has if policy 2 is followed is *strictly* smaller than the complaint that  $f(i)$  has if 1 is followed, or there are some individuals in  $S_2$  outside the range of  $f$ .
    - iii. Then one ought to follow policy 2, rather than 1.
  - b. Consequences (and non-consequences) of this method
    - i. One ought to save the greater number, in straightforward Taurekian cases.
    - ii. One ought to save B’s life and save C’s legs/cure C’s sore throat, rather than saving only A’s life.
    - iii. Cases in which this method delivers *no* judgment re. what one ought to do:
      - 1. The original Gladiator-style cases.
      - 2. Policy 1 saves one teenager from terminal cancer, but (for budget-balancing reasons) requires shutting down all the secondary schools, thus depriving millions of secondary education. Policy 2 is the status quo.

3. One can either (Policy 1) prolong A's life by 20 years, or (Policy 2) prolong B's life by 12 years and C's life by 8 years.
- iv. Worries arising from these non-consequences
  1. Its silence on these cases means that this criterion is implausible as a *necessary* condition for when a policy ought to be pursued. (And note that [for the appropriate notion of complaint-strength] the utilitarian/prioritarian too agrees that it is a *sufficient* condition.)
  2. The fact that the contractualists' intuitive judgments on the above cases are non-uniform suggests (?) that it will be difficult to find a non-aggregative way of supplementing the method that will allow it to deal in the way they want with all such cases simultaneously.
5. **Method 3:** The "virtual-divisibility approach" (Kamm, sec. VII)
  - a. When comparing policies A and B, again identify the beneficiary groups  $S_1$  and  $S_2$  as above.
  - b. Suppose there exists a way of *splitting up* the benefits that individuals in  $S_1$  (resp.  $S_2$ ) gain from 1 (resp. 2) into a collection of benefits  $B_1$  (resp.  $B_2$ ), such that there is a one-to-one function  $f$  from  $B_1$  into  $B_2$  with the above properties. Then one ought to pursue 2.
  - c. Consequences of this method
    - i. One ought to save the greater number, in Taurek-style cases (as before).
    - ii. One ought to save B's life and C's legs/sore throat, rather than only saving A's life (as before).
    - iii. One probably ought to prolong B's life by 12 years and C's life by 8 years, rather than only prolonging A's life by 20 years (for reasons of baseline life-length).
    - iv. Verdicts on the "teenage cancer" and Gladiator cases?...
  - d. Worry: This method is(?) just a rewriting of utilitarianism (or prioritarianism.) Relatedly: It altogether jettisons any principle of separateness-of-persons, so has given up on what was supposed to be distinctive about the "complaint"-based approach.
6. The only hope? Explicitly aggregative theories with some sort of "diminishing marginal returns" structure<sup>1</sup>
  - a. E.g. the "moderate trade-off theory" discussed by Carlson: The badness of an array of harms ( $x_1, \dots, x_n$ ) to distinct individuals, ordered so that  $x_1$  is the most and  $x_n$  the least severe harm (equivalently,  $x_1$  is the smallest benefit,  $x_n$  the largest benefit), is given by
 
$$\sum_i x_i k^{i-1}$$
 (where  $k < 1$ ).
  - b. The point of this theory is: A single sufficiently large harm (e.g. being a gladiator) will outweigh *any number* of sufficiently smaller harms (e.g. being deprived of watching a gladiator match).

<sup>1</sup> For miscellaneous related ideas in various contexts, see: Carlson, "Should we kill to avoid headaches?"; Binmore and Voorhoeve, "Defending transitivity against Zeno's paradox"; Broome, "No argument against the continuity of value: reply to Dorsey"; Sider, "Might Theory X be a theory of diminishing marginal value?".

- c. Concerns about this theory:
    - i. Its verdicts are sensitive to choice of status quo. (*As it stands* – at any rate – the method issues the opposite verdict in Gladiator cases if the status-quo is that the entertainment will go ahead.)
    - ii. Since it directly evaluates *profiles of benefits/harms*, rather than states of affairs, it is not a value function in the usual sense.
    - iii. It's not clear that these defects can be removed; the obvious ways of formulating analogous true value functions don't give results that *anyone* wanted. (Cf. the "geometrist" theory discussed by Sider.)
  - d. Open research question: Can this idea be used to ground any sane approach?
7. Conclusions
- a. We have failed (?) to find any *distinctively individualist* alternative to standard consequentialist aggregation (utilitarian or otherwise) that issues judgments that are at all plausible in all cases.
  - b. If there isn't any, the anti-utilitarian rhetoric of "separateness of persons" is just so much smoke and mirrors.

**Appendix: Others on the “separateness of persons”-objection to utilitarianism**

Nagel: Utilitarianism “ignores the distinction between persons... To sacrifice one individual life for another, or one individual’s happiness for another’s is very different from sacrificing one gratification for another within a *single* life.” (*The possibility of altruism*, p. 142)

Nozick: “Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm... In each case, some cost is borne for the sake of the greater overall good. Why not, *similarly*, hold that some persons have to bear some costs that benefit other persons more, for the sake of the overall social good? But there is no *social entity* with a good that sacrifices 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 the others. ... Talk of an overall social good covers this up. (Intentionally?) To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. *He* does not get some overbalancing good from his sacrifice.” (*Anarchy, state and utopia*, pp. 32-3)

Rawls: “Each member of society is thought to have an inviolability... which even the welfare of everyone else cannot override. Justice denies that the loss of freedom for some is made right by a greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded. ... [F]rom the standpoint of contract theory one cannot arrive at a principle of social choice merely by extending the principle of rational prudence to the system of desires constructed by the impartial spectator. To do this is not to take seriously the plurality and distinctness of individuals.” (*A theory of justice*, pp. 28-9)

Otsuka and Voorhoeve: “[A] shift of weighting when we move [from the intrapersonal case of risk] to the interpersonal case [where the question is one of distribution between people, rather than between states of nature for a single person] can be resisted only on pain of denying the moral significance of the separateness of persons. This is because a single person has a unity that renders it permissible to balance (expected) benefits and burdens against each other that might accrue to her. A group of different people, by contrast, does not possess such unity. As a consequence, some forms of balancing benefits and burdens that are permitted when these accrue to a single person are impermissible in cases where these benefits and burdens accrue to different people.” (“Why it matters that some are worse off than others”, p.179)

## **Value Theory Seminar: HT14 Week 7**

Readings marked “\*” are available on Weblearn.

### **Core reading (mandatory)**

This “core” reading list is longer than usual. If you have to skip something, it’s probably best to skip either Scanlon or Schroeder, but none is ideal. The way they fit together is: Scanlon outlines and motivates (*inter alia*) a fitting-attitude analysis of value; Rabinowicz and Ronnow-Rasmussen criticise that analysis on the basis of the “wrong kind of reasons” problem; Way and Schroeder explore/defend two different ways of solving that problem.

Scanlon, *What we owe to each other*, Chapter 2, sections 1-4.

\*Rabinowicz and Ronnow-Rasmussen, “The strike of the demon: On fitting pro-attitudes and value”

\*Way, “Transmission and the wrong kind of reasons”

\*Schroeder, “Value and the right kind of reasons”

### **Further reading (optional)**

SEP articles on “fitting attitude theories of value” (<http://plato.stanford.edu/entries/fitting-attitude-theories/>), “Value theory” section 3.2 (on “fitting attitudes”: <http://plato.stanford.edu/entries/value-theory/#FitAtt> ), and the associated “Supplement on three ways of solving the wrong kind of reasons problem” (<http://plato.stanford.edu/entries/value-theory/supplement3.html> ). Also references therein.

Ewing, *The definition of good*, ch. 5 (“An analysis of good in terms of ought”)

\*Bykvist, “No good fit: Why the fitting-attitude analysis of value fails”

\*Anderson, *Value in ethics and economics*, ch. 2 (“A pluralist theory of value”).

Parfit, *On what matters*, Volume 1, ch. 10 (“Respect and value”).

**Fitting-attitude theories of value**

1. Introduction: Buck-passing, fitting-attitudes, non-teleology
  - a. According to (the letter + spirit of) classical consequentialism, (i) one ought always to maximise the good, (ii) one ought to perform the actions in question *because* they maximise the good, and (iii) this exhausts the ways in which considerations of goodness/value plug into deontology.
  - b. Three (related?) deviations from the classical-consequentialist picture
    - i. *Non-teleologists* deny that the connection between goodness/value and deontology is exclusively a matter of *promoting* the good. E.g. The value of a painting might relate rather to the fact that one ought to *contemplate* and *admire* it. Scanlon's key examples: Friendship, science.
    - ii. *Buck-passing*: For any object that is good, there are certain natural properties of the object in virtue of which it is good, and because of which one ought to respond to the object in question in certain ways. But goodness is not a further first-order property playing any separate or additional role in the explanation of what one ought to do.
    - iii. *Fitting-attitude analysis of 'good'/'valuable'*: For an object X to be good/valuable *just is* for X to be such that there are reasons to/one ought to/it is fitting to have certain "pro-attitudes" (desiring to promote, admiring, desiring to study, 'respecting' etc.) towards X.
      1. (It's not clear what, if anything, hangs on the choice of formulation in terms of reasons/ought/fittingness.)
  - c. These three theses are *conceptually* distinct:
    - i. Buck-passing/FA vs teleology:
      1. One can clearly be a fitting-attitude theorist without denying teleology (by restricting the relevant "pro-attitudes" in one's FA account to promotion-related ones). The same thought shows how one can be a buck-passer without denying teleology.
      2. Conversely: there doesn't seem to be any reason why one couldn't in principle deny teleology without backing up that denial with any sort of fitting-attitude account.
    - ii. Buck-passing vs FA:
      1. Buck-passing without FA: Hedonists hold that being *pleasant* is the only good-making feature that a thing can have. A hedonist can easily 'pass the buck' from goodness to pleasure, without thereby incurring any obligation to *analyse* goodness in terms of reasons-to-respond or anything else.
      2. In contrast, an FA analysis of goodness/value does(?) entail 'passing the buck': if one holds that for X to be good *just is* for there to be reasons to respond positively to X, then one cannot consistently also hold that the property of goodness generates independent reasons of its own for positive response.
    - iii. It is, however, very natural to hold all three theses together (as Scanlon does). Scanlon's account (cf. esp. WWOTEO pp.95-7):

1. For X to be good/valuable just is for X to have first-order properties playing the value-conferring role (buck-passing).
  2. The value-conferring role is: being such that there are reasons to hold one or more “pro-attitudes” towards X (fitting-attitude analysis).
  3. The “pro-attitudes” in question can include not only attitudes related to promoting X, but also admiring, ‘respecting’, etc. (non-teleology).
2. What might turn on the question of whether or not an FA analysis is correct (an incomplete and preliminary sketch!)
    - a. The priority of the deontic over the evaluative: The spirit of classical consequentialism takes the evaluative to be explanatorily prior to the deontic. But if the evaluative can be *analysed in terms of* the deontic, this cannot be correct.
    - b. Teleology vs non-teleology
      - i. In Scanlon’s account, the FA analysis is a key part of the case for *rejecting* teleology.
      - ii. Several other authors (Dancy, Smith, Portmore<sup>1</sup>) have argued the precise opposite: that an FA analysis of value guarantees the truth of teleology (roughly, because given an FA analysis of good/value, anything there is most reason to promote automatically qualifies as best/most valuable, at least in the promotion-related sense of ‘best’/‘valuable’).
    - c. Population ethics: One argument/intuition-pump in favour of the ‘total utilitarian’ population axiology proceeds from the thought that *life is valuable* to the thought that other things being equal, *more life is better*. But this move relies on a teleological conception of value.
  3. Problems for an FA analysis of goodness/value
    - a. Circularity threat: Note that the notion of there being *reasons to hold* certain pro-attitudes/those pro-attitudes being *fitting* cannot themselves be analysed in terms of the attitudes’ being *valuable*, on pain of circularity (given the reductive ambitions of the analysis).
    - b. Unification problem: What unifies the various attitudes that are all supposed to be “pro”?
    - c. The “wrong kind of reasons” (WKR) problem: for any pro-attitude, there can be reasons to have that pro-attitude towards an object X that have nothing to do with the matter of whether or not X is valuable.
      - i. For example, if a demon threatens to kill me unless I desire a saucer of mud, that (apparently!) gives me a reason to desire the saucer of mud; but it does not follow that the saucer of mud is desirable/valuable. (Cf. also the “paradox of hedonism”/“self-defeatingness” objections to consequentialism.)
      - ii. Analogies to cases of intention and belief
        1. Intention: The “toxin puzzle”

---

<sup>1</sup> Dancy, “Should we pass the buck?”; Smith, “Two kinds of consequentialism”; Portmore, “Consequentialising moral theories”

2. Belief: Epistemic vs pragmatic reasons for belief; Pascal's Wager
- d. The "problem of solitary goods" (Bykvist): there are some good states of affairs that it is logically impossible to bring about (because the specification of the state of affairs excludes the existence of any agents who could do the bringing-about). It cannot be fitting to pursue (or, arguably, to desire) something that it is logically impossible to bring about. But the state of affairs in question could be good nonetheless.
4. Three basic strategies for solving/avoiding the WKR problem:
  - a. Provide a (non-circular) criterion for distinguishing the right from the wrong kind of reasons;
  - b. Deny that there really are any reasons of the wrong kind (WKR-skepticism);
  - c. Replace talk of reasons/ought/etc. with a basic deontic notion of fittingness.
    - i. First worry: Vicious circularity looming? (As above)
    - ii. Second worry: *If* all deontic notions must ultimately be understood in terms of reasons (?), then this only postpones the problem.
5. Solution-attempt #1: object- vs state-given reasons
  - a. Proposal: Reasons of the right kind (RKR) to (e.g.) desire X are "object-given", i.e. given by properties of X itself; reasons of the wrong kind (WKR) are "state-given", i.e. given by the state of desiring.
  - b. Objection: The required distinction between object-given and state-given reasons cannot be made sufficiently precise:
    - i. because of "Cambridge properties";
    - ii. and anyway, in examples like: a demon threatens us with death unless we admire *him*. Our reasons to admire this demon are equally given by "properties of the demon" and by "properties of the state of desiring the demon".
6. Schroeder's solution-attempt
  - a. Proposal: RKRs for engaging in activity A in a certain way are reasons that would be reasons for *anyone* who is engaging in A at all to do so in that way, and just *because* she is engaging in A. Examples:
    - i. A is the activity of laying the table for a White House dinner; we consider the "way" that involves placing the salad forks next to the soup spoons.
    - ii. A is the activity of forming beliefs as to whether or not John is innocent; the "way" under consideration is believing that John is innocent.
  - b. Nice feature of this account: It applies equally to any activity that is governed by standards of correctness, not only to activities that are the havings of attitudes. Since the RKR/WKR distinction also applies to that wider domain, one would expect a unified account of that distinction. Alternative accounts tend to apply only to attitudes, or even more narrowly to attitudes of certain types.
  - c. Worry: Schroeder's proposal for how to analyse good-*for* looks likely to lead to vicious circularity.
7. Way's (WKR-skeptical) solution-attempt
  - a. Proposal: So-called WKRs for having attitude A are really (RK) reasons for *wanting* to have A/*for trying to bring it about* that we have A. There are no WKRs.
  - b. Worries (R&R-R): This proposal

- i. is *ad hoc*;
  - ii. would anyway only shift the problem: we will then need an account of when something is a reason *for the attitude vs for trying to bring it about that we have that attitude*.
- c. Way's reply to the ad-hocness charge: There is independent argument for WKR-skepticism:
  - i. RKR and WKR "transmit" in different ways. Examples:
    - 1. (RKR pattern 1: Intention pattern) If there is a RKR to intend to A, then the fact that B-ing facilitates A-ing is a RKR to intend to B.
    - 2. (RKR pattern 2: Desire pattern) If there is a RKR to desire A, then the fact B facilitates A is a RKR to desire B.
      - a. WKRs don't fit either of these patterns.
    - 3. (WKR pattern) If there is a WKR for attitude A, then the fact that attitude B facilitates attitude A is a WKR for attitude B.
      - a. RKRs don't fit this pattern.
  - ii. The difference in transmission patterns calls for explanation.
  - iii. WKR-skepticism can explain the difference in patterns, via appeal to independently established patterns for reasons for action/reasons to bring about given states of affairs.
  - iv. Claim: WKR-defenders cannot explain the observed transmission patterns (usually: because they fail to say enough in *positive* terms about WKRs to ground any such explanation), except on pain of "reductionism" ("For p to be a reason of the wrong kind for attitude A is for p to be a reason of the right kind to want attitude A"). In turn, reductionism is to be rejected because reductionists cannot satisfactorily "show that their analysis entails that the wrong kind of reasons for an attitude really are reasons for that attitude."
- d. Nice feature of Way's account: This account treats incentives for attitudes in the same way as one would treat incentives for the sort of things that there *cannot* be reasons for, as e.g. when a demon threatens to kill you unless you have a headache.
- e. Lingering worry: Way hasn't answered R&R's charge of merely shifting the problem.