The normative-theoretic implications of reductionism about personal identity

Hilary Greaves, University of Oxford

Formal Ethics conference, University of York, June 2017

Reductionism about personal identity

- ‘Timeslice’: a primitive
- The question of personal identity: under what conditions is timeslice x *part of the same person as* person y?

  - Reductionist view: The question can be answered in terms of relations we have independent reason to postulate (memory, intention, similarity… “R-relations”)
  - ‘Further fact’ view: when a *sui generis* relation (personal identity!) holds between x and
  - Eliminativist view: never/the question suffers from presupposition failure
- Next question: What are the implications of reductionism for normative theories?

Outline

- 3 reactions to reductionism: 'directional', 'conservative', 'revisionist'
- Revisionism (I): Non-aggregative questions
- Revisionism (II): Prudential value
- Revisionism (III): Moral value
- Summary
- Some questions

The directional reaction

- 'Boundary-ignoring' normative theories: Those according to which the boundaries between persons have no normative significance [and neither does anything like them].
  - E.g. 'Complete utilitarianism'
- 'Boundary-respecting' normative theories: The rest.
  - E.g. lifetime prioritarianism/egalitarianism, theories of desert…
- First reaction: Reductionism increases the plausibility of boundary-ignoring vs. boundary-respecting theories.
  - Parfit (I): Because personal identity is a *less deep matter* on a reductionist view than on a further-fact view.
  - Objection: There is no obstacle to normative theories trading in concepts that are "non-deep" in this sense (i.e.(?), non-fundamental).
    - Clearly they do trade in at least some non-fundamental concepts (of mental states, etc.).

The conservative reaction

- Second reaction: Given reductionism, whether or not the personal identity relation holds will sometimes be indeterminate. Therefore, if a boundary-respecting normative theory is true, normative facts will sometimes be indeterminate. (*But* no other revisions are required.)
- Example: It's indeterminate whether or not Omega is Alpha. Is it morally permissible for Alpha to accept some benefit for which Omega must pay the cost? (Williams 2016a)
- But there are anyway plenty of other sources of normative indeterminacy. (Dougherty 2014, Schoenfield 2016, Williams 2016a, Williams 2016b)

The revisionist reaction

- Third reaction: Given reductionism, "personal identity is not what matters":
  - Consider e.g. fission cases. Fission must be at least as good as ordinary survival, in prudential terms. But Parent cannot be identical to either Lefty or Righty, nor identical to both. Therefore identity is not what prudentially matters. (Parfit 1984, sec. 90)
  - (And presumably the same arguments apply equally to other kinds of normativity.)
- Parfit (II): Normative theories should be **"naturally expressible in terms of"** the reduction basis, rather than the personal identity relation itself.
  - "R-relation": assigns numbers $R(x,y) \in [0,1]$ to ordered pairs of timeslices, representing the degree to which timeslice y is relevantly related to timeslice x
    - Graded
    - "Not an equivalence relation" (e.g. strong R-relatedness not transitive)
  - Project: Investigate what normative theories can then look like.

Strands of the revisionist program

- There are (at least) 3 things the revisionist needs to take on: What to say about
  - **(1)** Questions normally taken to depend on whether or not the personal-identity relation holds between 2 given timeslices. ('Non-aggregative questions.')
  - Aggregative questions
    - **(2)** of prudential value;
    - **(3)** of moral value.

The revisionist program (I): Non-aggregative normative questions

- Non-aggregative questions arise when we consider ethical issues that concern only the personal-identity-like relations between two particular timeslices. E.g.
  - Is it *morally permissible* for Alpha to accept some benefit for which Omega must pay the cost?
  - Is it *prudentially rational* for Alpha to accept some cost in order that Omega receive a benefit?
  - Does Omega *deserve punishment/blame/etc*. for actions taken by Alpha?
- The conservative's reaction: Given reductionism, it becomes indeterminate whether normative verdicts (of permissibility/deservingness/etc.) apply.
- The revisionist's reaction: The key concepts in these normative verdicts (permissibility, deservingness, etc.) must admit of *degrees*, corresponding to the degree (strength) of the R-relation.
  - These two reactions are very similar. (A merely terminological distinction?)

The revisionist program (II and III): Aggregative normative questions

- Things get more complicated here…
- Preview:
  - On an orthodox (non-reductionist) picture, the notion of *lifetime well-being* is key, for both prudential and moral purposes.
  - But the revisionist is not entitled to the notion of lifetime well-being. The question then arises of how to rewrite normative theories in revisionist-acceptable terms.
  - It's clear enough how to do this for the prudential case.
  - The moral case is more difficult. (Suggests an argument against revisionism.)
  - But – I will argue, on a case by case basis – it too can (probably) be done.
    - Prioritarianism
    - Egalitarianism
    - Sufficientarianism

An orthodox picture of prudential and moral value

- Prudential value:
  - Momentary well-being: how good state of affairs x is for person S at time t
    - Apply substantive theories of well-being at this level (probably)
  - Lifetime well-being: How good state of affairs x is for person S (considering S's life as a whole)
    - Source of prudential reasons for the agent
- Moral value ("from the point of view of the universe")
  - Assuming welfarism (or "when other things are equal" etc.), this is determined by aggregating lifetime well-being levels (according to a utilitarian/prioritarian/egalitarian/etc. aggregation function)
- But the revisionist's strictures block use of any notion of *lifetime* well-being.

Revisionism and prudential value

- Obvious way to formulate a reduction-basis theory of *prudential* value:
  - Prudential value of state of affairs A for timeslice x:

    $PV_x(A) = \sum_{y \in T} R(x,y) w(y,A)$, where

    $w(y,A)$ is the timeslice well-being level of timeslice y in state of affairs A;

    R(x,y) is the strength of the R-relation between x and y

    T(A) is the set of timeslices (assume fixed)
- Some controversial (but not implausible) implications
  - Discounting 'future' well-being is prudentially rational
  - Weak prudential reason to care about what happens to oneself following radical personality change/memory loss/etc.
  - Nonzero prudential reason to care about what happens to (some) other people

Revisionism and moral value: "The simple theory"

- On the orthodox account, to construct an account of moral value, we aggregate (i) the quantities that govern prudential rationality (ii) across the subjects of prudential rationality.
  - This suggests (in the present context) aggregating *prudential value* across *timeslices*.
- We can't aggregate across only *present* timeslices, on pain of time-relativity.
  - Fix: aggregate across *all* time-slices.
- Simplest way to do this ("**the simple theory** [of prudence-aggregated moral value]"):
  - $MV(A) = \sum_{x \in T} PV_x(A)$,

  where $MV(A)$ is the moral value of state of affairs A, and T is the set of timeslices that exist in A.

Problem with the simple theory: Multiple-counting

- Problem: The simple theory implies that a given episode of happiness/suffering matters less, the fewer/weaker the R-relations that the corresponding timeslice bears to other timeslices. This implication is not plausible.
  - Example: Preventing Annie's vs. Bob's 40^{th} birthday pain
- Some analogies
  - What utilitarianism (or any other standard theory of aggregation) says about the flourishing of cared-for children vs lonely orphans
  - What Scheffler (1994) says about the role of the 'personal point of view'
- The general moral: 'damped partial sums' can play various *other* roles in normative theory, but don't aggregate *those* things when seeking an account of agent-neutral moral value (on pain of multiple-counting).
  - Special case: Aggregating prudential value across time-slices is a terrible idea.

A new argument for ‘directionalism’?

- An argument:
  - (1) Parfit’s argument against conservatism is sound.
  - (2) But revisionism just doesn’t work, at the level of detail.
  - (3) Therefore the right reaction to reductionism must be the ‘directional’ one, i.e. reductionism should lead us to accept boundary-ignoring normative theories.
- What (the rest of) this talk is doing: Weakening the case for (2).

Revisionism-compatible theories of moral value

- The simple theory (and any distribution-sensitive variant on it) runs into problems because it tries to *aggregate* prudential values across timeslices.
- But this doesn't mean a sensible theory can't *somehow use* information about timeslices' prudential values.
- An alternative approach: *Aggregate* momentary well-being, but "apply distribution-sensitivity modifications via prudential value".

Prioritarianism

- Basic idea of prioritarianism: "Well-being has diminishing marginal moral value"
- Standard formalisation of this:

$$MV(A) = \sum_i f\big(w_i(A)\big)$$

where $f(0) = 0$, *and* f is strictly increasing and strictly concave.

Heuristic interpretation: $f(w)$ is the moral value of well-being level w.

- Two interpretations
  - Timeslice prioritarianism: i indexes timeslices; w_i is i's momentary well-being.
  - Lifetime prioritarianism: i indexes persons; w_i is i's lifetime well-being.

A problem for timeslice prioritarianism: no notion of 'compensation'

- (I.e.,) for the purpose of evaluating benefits at t, timeslice prioritarianism takes no account of differences in persons' well-being levels at any times other than t.
- Example (Holtug (2010):

		Momentary well-being at t_1	Momentary well-being at t_2	Lifetime well-being
"Status quo"	a	2	2	4
	b	4	2	6
"Benefit a"	a	2	4	6
	b	4	2	6
"Benefit b"	a	2	2	4
	b	4	4	8

- Timeslice prioritarianism is indifferent between "Benefit a" and "Benefit b".
- But intuitively, a prioritarian should judge "Benefit a" better.

A problem for lifetime prioritarianism: inappropriate ‘compensation’

- Lifetime prioritarianism takes account of persons’ well-being levels at other times *even when this seems inappropriate*, due to long times and weak connections. Example:

		Momentary well-being at t_1	Momentary well-being at t_2	Momentary well-being at t_3	Lifetime well-being
“Status quo”	c	4	8	4	16
	d	8	8	2	18
“Benefit c”	c	4	8	6	18
	d	8	8	2	18
“Benefit d”	c	4	8	4	16
	d	8	8	4	20

- Lifetime prioritarianism judges “Benefit c” better than “Benefit d”.
- But if the time-lag between t_1 and t_3 is very large, this seems inappropriate (at least by reductionist lights!).

Prioritarianism for revisionists

- Note that the standard prioritarian formula can equivalently be written
  - $MV(A) = \sum_i g\big(w_i(A)\big)w_i(A)$,

  where $g(w) \coloneqq \frac{f(w)}{w}$, *f* as above.

  Heuristic interpretation: $g(w)$ is the "priority weighting" at the well-being level w.
- What the problems of no/inappropriate 'compensation' suggest is that the priority weight (for a given timeslice)
  - *should depend* on what happens in the same life at other times, but
  - only *when/insofar as* there is a strong R-relation to the 'other time' in question.
- This naturally suggests basing priority weights on timeslices' prudential values:
  - $MV(A) = \sum_{x \in T} g\big(PV_x(A)\big)w_x(A)$.
  - "Benefitting timeslices matters more, the lower their prudential value."
  - Call this theory of goodness "R-prioritarianism". (Thomas (personal communication); *somewhat* similar to Holtug's "prudential prioritarianism" (2010, sec. 10.6).)

R-prioritarianism and the problems of no/inappropriate compensation

- Suppose that
  - The R-relation between adjacent time-slices is always ½;
  - The R-relation between corresponding timeslices at t1 vs. t3 is always ¼.
- Then R-prioritarianism gives the prioritarian-intuitively desired verdicts on the above examples:

		Momentary well-being at t_1	Momentary well-being (PV) at t_2	Lifetime well-being
"Status quo"	a	2	2 **(3)**	4
	b	4	2 **(4)**	6

Better to benefit a than b (at t_2)

		Momentary well-being at t_1	Momentary well-being at t_2	Momentary well-being (PV) at t_3	Lifetime well-being
"Status quo"	c	4	8	4 **(9)**	16
	d	8	8	2 **(8)**	18

Better to benefit d than c (at t_3)

Egalitarianism

- Basic idea of egalitarianism: "Inequality is intrinsically bad"
- Standard formalisation of this:

$$MV(A) = \left(\sum_i w_i(A)\right)(1 - I(\{w_i(A)\}_i)),$$

where

$\sum_i w_i(A)$ is total well-being in A,

I is a measure of "amount of inequality".

- Two interpretations…
  - Timeslice egalitarianism: A is an *instantaneous* state of affairs, at some given time t. i indexes timeslices existing at t. w_i is i's momentary well-being.
  - Lifetime egalitarianism: A is a timeless state of affairs. i indexes persons. w_i is i's lifetime well-being.

R-egalitarianism

- Timeslice and lifetime egalitarianism face the same problems of "no compensation" and "inappropriate compensation" (respectively) as the corresponding versions of prioritarianism.
- Natural suggestion: What the egalitarian should care about is inequality *of prudential values* among timeslices. Formally, perhaps
  - $MV(A) = \left(\sum_i w_i(A)\right)(1 - I(\{PV_i(A)\}_i))$.
- Like R-prioritarianism, this theory also gives the desired verdicts on the problems of no/inappropriate compensation.

Sufficientarianism

- Basic idea of sufficientarianism: "Everyone (or as many as possible) should have enough"
- Standard versions
  - Timeslice sufficientarianism: Maximise the number [measure] of timeslices whose momentary well-being is above the sufficiency threshold
  - Lifetime sufficientarianism: Maximise the number of people whose lifetime well-being is above the sufficiency threshold
- These views will give rise to a similar dilemma s in the prioritarian/egalitarian cases. Thus suggesting:
- R-sufficientarianism: Maximise the number [measure] of timeslices whose prudential value is above the sufficiency threshold.

Summary

- ‘Revisionism’: Rewrite/replace ‘boundary-respecting’ normative theories so that they make use only of (and can be ‘naturally expressible in terms of’) R-relations, rather than personal identity *per se*.
- This program is easy for non-aggregative and prudential questions.
- Less obvious whether the program is viable for aggregative questions of moral value.
  - A new argument for directionalism?
- This strikes me as too pessimistic. ‘R-prioritarianism’ and its (egalitarian/sufficientarian/etc.) analogues are worth further scrutiny.

Some more questions

- Worries about making so much use of 'timeslices'. (Cannot be subjects of well-being? Cannot be agents?)
  - But 'time-segments' seems an unstable compromise. (?)
  - And using 'persons relative to times' would give up the game.
- Holtug's (2010) prioritarian theory is similar to the 'R-prioritarian' theory above, but also places value on strengthening R-relations (assuming momentary well-being levels are positive).
  - Is this a thing we (/the prioritarian) should want? If so, how to formulate the desired theory (formally)?
  - Similar questions for egalitarianism/sufficientarianism.
  - (Population-axiology problems loom. ~Thomas (MS).)
- Precisify the *motivation* for revisionism. (No plausible *general* principle to the effect that one should rewrite normative theory in terms of the reduction basis…)

References

- Dougherty, T. (2014). Vague Value. *Philosophy and Phenomenological Research*, *89*(2), 352-372.
- Holtug, N. (2010) *Persons, interests and justice*. OUP.
- Parfit, D. (1984) *Reasons and Persons*. OUP.
- Scheffler, S. (1994) *The rejection of consequentialism*. OUP.
- Schoenfield, M. (2016). Moral Vagueness Is Ontic Vagueness. *Ethics*, *126*(2), 257-282.
- Thomas, T. (MS) Time-Relative Interests and Population Ethics.
- Williams, J. R. G. (2016). Indeterminacy, Angst and Conflicting Values. *Ratio*, *29*(4), 412-433.
- Williams, J R G (2016b). Indeterminate oughts. *Ethics 127(3):645-673*.