

# Well-Being for Changing Selves

Krister Bykvist  
Jesus College, Oxford

# Plan

- Introduce a challenge for all desire theories and endorsement theories, the challenge of changing desires and preferences.
- Present a framework for representing changing desires and preferences.
- Show that the challenge is more serious than is usually thought: threat of incoherence!
- Argue for a particular solution to this challenge. Focus on some recent objections.

# Desire theory

- (1)  $x$  is intrinsically *good* for you iff you intrinsically want  $x$ .
- (2)  $x$  is intrinsically *better* for you than  $y$  iff you intrinsically prefer  $x$  to  $y$ .

# Necessary refinement

- wanting x' should not be understood as 'preferring x to its negation'.
- Satisfying your want that p is not good in itself for you, if you take a *neutral* attitude towards p and a negative attitude towards not-p.
- Example: want not to have a headache

# Polarity of attitudes

- To *favour*  $x$  = to be positively oriented towards  $x$  in your actions, emotions, feelings or evaluative responses.
- To *disfavour*  $x$  = to be negatively oriented towards  $x$  in your actions, emotions, feelings or evaluative responses.
- To be *neutral* (indifferent) towards  $x$  = to be indifferently oriented towards  $x$  in actions, emotions, feelings, or evaluative responses.

# Polarity-based desire theory

- $x$  is good for you iff you favour  $x$ .
- $x$  is bad iff you disfavour  $x$ .
- $x$  is neutral for you iff you are neutral towards  $x$ .
- $x$  is better for you than  $y$  iff you prefer  $x$  to  $y$ .

Note : no need to treat absolute attitudes as primitive.

Favouring, disfavouring, and indifference can be reduced to comparative attitudes. (Chisholm)

- You favours  $x$  iff you prefer  $x$  to something you are indifferent towards.
- You disfavour  $x$  iff you prefer  $y$  to  $x$  and  $y$  is something you are indifferent towards.
- You are indifferent towards  $x$  iff you are indifferent between  $x$  and the negation of  $x$ .

# Objections to desire theory

Worthless objects of desires:

- Rawls' person who wants to count the blades of grass in collage lawns.
- Kraut's boy who wants to throw a rock at a duck.
- Quinn's person who has an irresistible urge to turn on radios.
- Sen's hopeless beggar, precarious landless labourer, dominated housewife, hardened unemployed, and over-exhausted coolie who all take pleasures in small mercies.

Satisfying these wants is not good in itself for these persons, (or, at least not good to a degree proportionate to the strength of the wants).



# Endorsement theories

Well-being consists in subjective endorsement of objective value.

Proponents: Adams, Darwall, Dworkin, Kraut, Olsaretti, Parfit.

## Endorsement constraints

### Absolute

Strict: x is good for you only if you favour x.

Weak: x is very good for you only if you favour x.

### Comparative

Strict: x is better for you than y only if you prefer x to y.

Weak: x is *much* better for you than y only if you prefer x to y.

# The challenge of changing preferences

How do we determine the well-being of a person when her preferences are not stable across worlds?

Examples:

*The career choice.*

If you become a philosopher, you will come to prefer this life to being a fiddler. If you become a fiddler, you will come to prefer this life to being a philosopher.

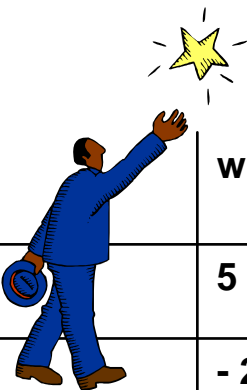
*The unmarried person's dilemma: 'To wed or not to wed'.*

If you get married, you will prefer not to be married. If you stay unmarried, you will prefer being married to being unmarried.

Which life is better for you?

The problem is to find a stable standard of well-being when the standard is set by preferences that are not stable.

## Toy model



	w1	w2	w3	..
w1	5	8	6	..
w2	- 2	-5	0	..
w3	4	4	10	..
:	..	..	..	..

- A row: representation of the attitudes the person has, in a certain world, towards her possible lives, greater numbers represent more preferred worlds, equal numbers represent indifference.
- A column: representation of all her possible attitudes towards her life in a certain world.
- Positive numbers represent favourings, negative numbers disfavourings, and zero neutral attitudes.

# Idealizing assumptions

- No intertemporal conflicts of attitudes.
- Same duration of lives.
- Global attitudes.
- Fully-opiniated preferrers.
- Properly laundered attitudes: informed, rational (preferences in a world are transitive, symmetric, and connected), autonomous etc.

# Preference reversal cases

## ‘The Career Choice’

	W1 (philosophy)	W2 (fiddle)	..
W1 (philosophy)	a	b	
W2 (fiddle)	c	d	
:			

- $a > b, d > c$ .

More information:

Would you *love* or *hate* your life as a fiddler?

Would you *love* or *hate* your married life?

## ‘To wed or not to wed’

	W1 (married)	W2 (unmarried)	..
W1 (married)	a	b	
W2 (unmarried)	c	d	
:			

- $b > a, c > d$ .

More information:

Would you *love* or *hate* your married life?

Would you *love* or *hate* your unmarried life?

# Assumptions that generate inconsistency

## *World-Bound Well-Being*

Your life in  $w$  is good for you iff you favour, in  $w$ , your life in  $w$ .

Some endorsement theories accept it only if it is restricted to lives that are objectively desirable or worthy of concern.

## *Comparative Endorsement*

Your life in  $w$  is better for you than your life in  $w'$  only if you prefer, in either  $w$  or  $w'$ , your life in  $w$  to your life in  $w'$ .

# Inconsistency

	w1	w2	..
w1	0	-2	
w2	8	6	
:			

***World-Bound Well-Being*** implies

(1) w1 is not good for you (since you do not favour, in w1, your life in w1)

and

(2) w2 is good for you (since you favour, in w2, your life in w2 and your life is objectively desirable)

So,

(3) w1 is either bad or neutral for you

So,

(4) w2 is better for you than w1 (since what is good must be better than what is bad or neutral)

But ***Comparative Endorsement*** implies

(5) w2 is not better for you than w1 (since in neither w1 nor w2 do you prefer w2 to w1)

So,

(6) w2 is both better and not better for you than w1.

**Contradiction!**



# Possible approaches

- a) Ideal desires
- b) Actual desires
- c) Relativism: good for S relative to a world w
- d) Think vertically
- e) Think horizontally
- f) Think diagonally

I will discuss a), e), and f) today.

# Evade the problem?

Define well-being in terms *ideal* desires and preferences, the ones we would have if we were in an epistemically ideal situation.

Crucial assumption: ideal desires and preferences cannot change.

*False*, if ideal desires depend in part on our actual psychological make-up.

Note that ideal attitudes need not be conditional on their own persistence:

I can favour x even though I know that I would not favour x if x obtained. Example: personal ideals.

*True*, if ideal desires are simply defined as desires that track objective desirability. But then desires become an idle wheel.

# Comparative Endorsement is counterintuitive

w2 cannot be better for you than w1 (no matter how much you would hate w1, if w1 obtained, and would love w2, if w2 obtained)

	w1	w2	..
w1	-20	-20	
w2	20	20	
:			

## My solution

Reject Comparative Endorsement. Stick to World-Bound Well-Being.

Bracketing the issue of objective desirability, we can also say:

- Your life in  $w$  is good for you iff you favour, in  $w$ , your life in  $w$ .
- Your life in  $w$  is bad for you iff you disfavour, in  $w$ , your life in  $w$ .
- Your life in  $w$  is neutral for you iff you are neutral, in  $w$ , towards your life in  $w$ .

# Think diagonally

Decide cross-world comparisons of well-being by looking at the values in the *diagonal*.

Your life in  $w$  is better for you than your life in  $w'$  iff

(i) You would *favour your life in  $w$  more*, if  $w$  obtained, than you would *favour your life in  $w'$* , if  $w'$  obtained,

or

(ii) You would *disfavour your life in  $w$  less*, if  $w$  obtained, than you would *disfavour your life in  $w'$* , if  $w'$  obtained,

or

(iii) You would *favour your life in  $w$* , if  $w$  obtained, and you would *disfavour your life in  $w'$* , if  $w'$  obtained,

or

(iv) You would *favour your life in  $w$* , if  $w$  obtained, and you would be *indifferent* towards your *life in  $w'$* , if  $w'$  obtained,

or

(v) you would be *indifferent* towards your *life in  $w$* , if  $w$  obtained, and you would *disfavour your life in  $w'$* , if  $w'$  obtained.

# Modified comparative endorsement

$w$  is better for you than  $w'$  only if your  
 $w$ -self *would* want  $w$  more than your  
 $w'$ -self *would* want  $w'$ .

# Application

	W1 (philosophy)	W2 (fiddle)	..
W1 (philosophy)	a	b	
W2 (fiddle)	c	d	
:			

- $w_1$  is better for you than  $w_2$  iff  $a > d$ , equally as good as  $w_2$  iff  $a = d$ .
- So if you *hate* your life as a philosopher and *love* your life as a fiddler, the life as a fiddler is better for you.

# No need to find a single vantage point

- Evades the problem of comparing very different lives from one single vantage point.
- Having the experiences necessary to appreciate what one kind of life is like may distort one's appreciation of what an alternative life would be like.
- Example: life as an Amish versus life in the city.



# Objections

(1) What about regret?

I would have to say that w1 and w2 have the same value for you, but w2 seems worse in one respect: you will have regrets, 'what is' compares unfavourably with 'what might have been' (Sugden)

	w1	w2	..
w1	5	2	
w2	10	5	
:			

# Replies to objection from regret

(1) How does the regret factor affect well-being?

Suggestion:

How well off I am in a world depends not only on what I feel about my life in that world but also how much I regret not living an alternative life.

A simple version: the well-being of a life in a world  $w$  is a function of

- (a) the intensity of the absolute attitude in  $w$  towards the life in  $w$  and
- (b) the regret factor: the difference between the intensity of the attitude in  $w$  towards the life in  $w$  and the intensity of the attitude in  $w$  towards *the highest ranked alternative*.
- The regret-factor is zero, if there is no alternative that is ranked higher.

Problem for this regret-sensitive view: Considering new alternatives might change the relative values of the original alternatives.

In Case 1, it seems plausible to say that w2 is better for you than w1 .

But the regret-sensitive theory does not guarantee such a conclusion

In Case 2, the regret-factor for w1 will be -20, whereas for w2 it will be -90. This means that in one respect your life in w2 is worse.

Case 1	w1	w2	..
w1	-10	10	
w2	-10	10	
:			

Case 2	w1	w2	w3	.
w1	-10	10	0	.
w2	-10	10	100	
:				

## Pairwise regret

How well off I am in a world  $w$  compared to another  $w'$  depends not only on what I feel about my life in each world but also how much I would regret living in  $w$  rather than  $w'$  or living in  $w'$  rather than  $w$ .

To compare  $w$  and  $w'$  we need to look at

- (a) the intensity of the absolute attitude in  $w$  towards  $w$
- (b) the intensity of absolute attitude in  $w'$  towards  $w'$ .
- (b) the regret factor for  $w$ : the difference between the intensity of the attitude in  $w$  towards  $w$  and the intensity of the attitude in  $w$  towards  $w'$ .
- The regret factor for  $w'$  in relation to  $w$ : the difference between the intensity of the attitude in  $w'$  towards  $w'$  and the intensity of the attitude in  $w'$  towards  $w$ .
- The regret-factor is zero, if the alternative is not ranked higher.

# Problem: Cyclical betterness

	w1	w2	w3
w1	5	2	10
w2	10	5	2
w3	2	10	5

- w1 is better for you than w2, since w1 comes with less pairwise regret (0 instead of -5)
- w2 is better for you than w3, since w2 comes with less pairwise regret (0 instead of -5)
- w3 is better for you than w1, since it comes with less pairwise regret (0 instead of -5)

(2) The regret-factor as it is defined seems  
*unnecessary.*

- Suppose that I am satisfied with my actual career, but feel deep regret that I was never able to write a certain book. The fact that I feel regret seems relevant to my well-being.
- **But** recall that the attitudes I am focusing on are global, about my life as a whole. We need to know what I think about having the career *while feeling deep regret*. Global attitudes will be sensitive to feelings of regret.

# Badness of frustrated comparative preference

	w1	w2	..
w1	5	2	
w2	10	5	
:			

- What makes w2 a worse alternative is the fact that in w2 my comparative preference for w1 over w2 is frustrated.
- Objection: having a frustrated preference is not bad in itself for you. It all depends on your attitude towards the less preferred state of affairs.

# Badness of frustrated favouring

- If satisfied favourings make my life better, shouldn't *frustrated* favourings make it worse?
- If 'yes', then we have a reason to rank  $w_2$  over  $w_1$ , because the intensity of my frustrated favouring in  $w_2$  is stronger than the intensity of my frustrated favouring in  $w_1$ .

Problem:

- Frustrations of favourings concerning suboptimal lives make a life worse.
- It is true that by leading a blissful life I miss out on other lives that are good in my own light. But the absence of a good does not make a bad.



Objection : My account presupposes cross-world intrapersonal comparisons of attitudes.

What does it mean to say my x-self wants x more than my y-self wants y?

Unproblematic cases (different polarity):

My x-self *favours* x, my y-self *disfavours* y.

My x-self *favours* x, my y-self is *neutral* towards y.

My x-self is *neutral* towards x, my y-self *disfavours* y.

But what does it mean to say that my x-self *favours* x *more* than my y-self *favours* y.

My x-self's preference for x over something he is indifferent towards is stronger than my y-self's preference for y over something he is indifferent towards.

- Comparisons of favourings = comparisons of preference *differences*.