

Utilitarianism by Way of Preference Change?

Making sense of Hare's argument

Moral Thinking, 1981

Main idea: Moral deliberation – a thought-experimentation process that results in preference change.

Several well-known problems with the argument.

Here: **The No-Conflict Problem**

A solution in Rabinowicz & Strömberg 1996.

Not quite satisfactory.

A new solution - Daniel Elstein

Alternative: To *dissolve* the problem - Vendler's move (1988)

BUT: This gives rise to a potentially more serious difficulty.

MORAL JUDGMENTS:

Overriding Universalizable Prescriptions (OUP)

Universalizability \Rightarrow Need for *thought-experiments* in moral deliberation:

- What if I were in his shoes?
- What if I were on the receiving end?

“Being in his shoes” – stringently interpreted:
Not only being in his external circumstances,
but also having his psychological make-up,
including his preferences.

EXAMPLE: My students, John and Mary, are waiting for me at the department.

Is it ok that I go there by bike, as I prefer,
or should I go by car, as they prefer?

Preferences of the persons involved - for or against going by bike

mine: +4

John's: -3

Mary's: -2

Signs: direction of preference Values: preference strength

Possibility of interpersonal preference measurement may be questioned, but here it is taken for granted.

Preference utilitarianism:

The balance of preferences tells against the bike alternative.

How can Hare arrive to the same result?

The first step in moral deliberation:

I envisage situations in which I occupy the positions of each of the persons involved.

The three situations and my preferences in these situations:

My actual situation:	s_1	+4
If I were in John's shoes:	s_2	-3
If I were in Mary's shoes:	s_3	-2

The second step:

I acquire preferences with regard to hypothetical situations.

CONDITIONAL REFLECTION (**CR**):

If I (fully) know what I would prefer *in* a hypothetical case, I must have the same preference (equal strength, equal sign) *with regard to* that case.

Hare: **CR** is a *conceptual truth*, due to the prescriptive element in the concept of “I”.

Insofar as I think of hypothetical preferences as *mine*, I thereby endorse them.

Alternative: **CR** might be taken as a *rationality requirement*. On this reading, **CR** characterizes ideally self-integrated preferrers.

Objection: What if my hypothetical preferences are seen by me as corrupt, from my current point of view?

But: In our example this problem doesn’t arise.

CR \Rightarrow I acquire several preferences concerning the bike alternative, one preference with regard to each situation:

- | | |
|----------------|----------------------------------|
| w.r.t. s_1 : | For biking, with strength +4 |
| w.r.t. s_2 : | Against biking, with strength -3 |
| w.r.t. s_3 : | Against biking, with strength -2 |

Third step: Preference balancing

Hare: The original *interpersonal* preference conflict is transformed into an *intrapersonal* one.

This balancing tells against the bike alternative – the same result as in preference utilitarianism.

The No-Conflict Problem

(G. F. Schueler 1984; I. Persson 1989)

The preferences acquired in accordance with **CR** concern *different* situations.

⇒ They do not oppose each other.

Thus, no intrapersonal preference conflict

⇒ No need for balancing.

But haven't we forgotten *universalizability*?

U comes in not only at the initial stage – as the basis for thought-experiments,
but also later – as a requirement on the output:

My prescription with regard to s_1 , s_2 and s_3 must be *the same* in order to be moral.

Geometrical solution (Rabinowicz & Strömberg 96)

CR \Rightarrow
My preference profile: $\begin{matrix} s_1 & s_2 & s_3 \\ (+4, & -3, & -2) \end{matrix}$

From this point of departure we need to arrive at a universal prescription – the same for all three situations.

Hare: Accepting a prescription = preferring.

Therefore: The task is to revise one's preferences so as to reach a uniform preference state:

$$(+4, -3, -2) \Rightarrow (x, x, x)$$

Question: $x = ?$

Revision of *preferences* can be compared with revision of *beliefs*.

PRINCIPLE OF MINIMAL CHANGE:

The output state should minimally differ from the input state.

Task:

Move from a given state to the closest uniform state (x, \dots, x) .

Determination of distance

A preference state may be seen as a point in a vector space. If the space is Euclidean, we get a measure of distance between preference states.

Euclidean distance is minimized iff x is the *average* of the values in the original pref. state.

My original state: $(+4, -3, -2)$

My revised uniform state: $(-1/3, -1/3, -1/3)$

Averaging: a preference utilitarian solution.

Questions about distance measure:

Why Euclidean?

Why not some other measure?

Would we then still get the preference utilitarian solution?

What are the adequacy criteria for an appropriate distance measure?

Suppose we want to determine the distance between two points:

$$(v_1, \dots, v_n) \text{ and } (w_1, \dots, w_n)$$

The simplest distance measure:

City-block: $\sum_{i=1, \dots, n} |v_i - w_i|$

City block: A non-utilitarian solution

- No unique solution in 2-person cases.
Ex. For input (3, -1) all solutions from (3,3) to (-1,-1) are equally acceptable
- When the solution *is* unique, it need not be utilitarian.
In those cases city block picks out the *median*.
Ex. For the input (6, 0, -3), averaging \Rightarrow (1, 1, 1),
while city-block \Rightarrow (0, 0, 0).

Minkowski measures:

$$[\sum_{i=1, \dots, n} |v_i - w_i|^k]^{1/k} \quad (k \geq 1)$$

$k = 1$ - city block, $k = 2$ - Euclidean distance, ...

$k > 1$ – Larger differences $|v_i - w_i|$ get disproportionately greater influence, as compared with smaller differences.

This looks like a consideration of *fairness*.

Utilitarianism is not supposed to give fairness any weight.

BUT: City block ($k = 1$) doesn't yield utilitarianism.
Only the Euclidean measure ($k > 1$) does.

Puzzle!

Questions about Harean exegesis:

How faithful is the geometrical solution to Hare's argument?

- The idea of minimal preference change is not exploited by Hare
- Hare's suggestion that interpersonal preference conflicts get transformed into intrapersonal ones is not made use of.

Simultaneous extrapolation (Daniel Elstein)

CR \Rightarrow I acquire preferences concerning the action with regard to each of $s_1 - s_3$:

(+4, -3, -2)

Next step:

U \Rightarrow I simultaneously extrapolate each preference in my preference state to all exactly similar situations $s_1 - s_3$:

($\langle +4, -3, -2 \rangle$, $\langle +4, -3, -2 \rangle$, $\langle +4, -3, -2 \rangle$)
 s_1 s_2 s_3

In this new preference state, I accept three universal prescriptions,
one *for* the action (strength 4) and two *against* (strengths 3 and 2).

But: How can I simultaneously accept incompatible prescriptions?

Answer: These prescriptions are only *pro tanto*.

Each reflects just one relevant aspect of the case.

Going by bike is universally prescribed *insofar as* I prefer it in s_1 .

It is universally forbidden *insofar as* I disprefer it in s_2 .

And so on.

Pro tanto prescriptions are universal, but not overriding.

Novelty:

Universalizability is imposed at an earlier stage of deliberation, on prescriptions that aren't yet *all-things-considered*.

Last step of moral deliberation:

In the state

$$\begin{array}{ccc} (<+4, -3, -2>, <+4, -3, -2>, <+4, -3, -2>) \\ s_1 & s_2 & s_3 \end{array}$$

my preferences conflict with regard to each of $s_1 - s_3$.

This *intrapersonal* preference conflict solved by balancing:

$$+4 - 3 - 2 = -1$$

I end up with the same preference all-told w. r. t. each of $s_1 - s_3$:

$$(-1, -1, -1)$$

↑↑ Overriding universal prescription, with the right utilitarian flavour.

How faithful is simultaneous extrapolation to Hare?

Quite faithful.

Main difference: Introduction of *pro tanto* moral oughts

Overridable *prima facie* oughts are recognised by Hare - general principles accepted at intuitive moral level admit of exceptions.

But overridable *pro tanto* moral oughts do not figure in his theory, as far as I know.

Zeno Vendler (*Hare and Critics*, 1988):

Contrary to what Hare suggests, his thought-experiments do *not* involve imagining new situations.

What is envisioned is still the *actual* situation,
but now viewed from someone else's perspective:

... there *seem* to be ... different situations envisioned ... Hare says this, 'Note that although the two situations are different, they differ only in what *individuals* occupy the two roles; their *universal* properties are all the same' (*MT* 111). 'No', I say, it is the same situation, with the same individuals; the only difference is which of them is I: in imagining being he, I imagine the same situation from a different perspective. (Vendler, p. 178)

What is being 'moved' from one role to another is not the person I am,
but the "transcendental I" – “the empty frame of consciousness”.

If Vendler is right, then:

(i) The No-Conflict Problem disappears

The different preferences I acquire (in accordance with **CR**) concern the same situation.

Therefore, they *can* conflict with each other.

However, if Vendler is right, then:

(ii) A new problem instead

CR is an expression of our fundamental *self-concern*.

This self-concern is extendable to hypothetical situations.

But: what "self" am I concerned about in this fundamental way?

The person I am, or the empty frame of consciousness?

The real I or the transcendental I?

Surely, the former.

But then: If Vendler is right, Hare's thought-experiments are outside CR's domain of application.

Thus: Preferences belonging to different perspectives remain uncollected.

Hare's argument never gets off the ground.

Different ways of dealing with the No-Conflict Problem

- Geometrical solution (Rabinowicz & Strömberg)
 - the choice of a distance measure must be justified
- Simultaneous extrapolation (Elstein)
 - pretty neat
- Dissolution of No-Conflict Problem (Vendler)
 - **CR** fails