Preference Change in the Multiple-Self

Conrad Heilmann

Department of Philosophy, Logic and Scientific Method
London School of Economics (LSE)

Preference Change Workshop, LSE Choice Group
28 May 2009

Summary

- The Multiple-Self Models as foundation for theories of preference change
- Application: Intertemporal Weakness of the Will

Preference Change and Time

- In standard decision theory, there is a entirely coherent and stable decision-maker, characterized by a set of preferences A.
- Considering preference change, the decision-maker will have at least two distinct sets of preferences such that $A = \{A_0, A_1\}$, where the index most commonly refers to time.
- That is, there are at two distinct instances in time in which a decision-maker is governed by a different set of preferences.
- More generally, when considering preference change, a decision-maker can be seen as $A = \{A_0, A_1, \ldots, A_k\}$, where k is a time index.

Preference Conflict at a Time

- However, the time dimension need not be invoked necessarily, as there can also be preference conflict at a time
- For instance, differing preferences due to some other dimension than the temporal one, for instance in moral conflicts or due to different evaluations
- More generally, when considering preference conflict, a decision-maker can be seen as a decision-maker with more than one set of preferences at a time: $A = \left\{A^0, A^1, \ldots, A^n\right\}$, where n is an index that refers to different situations or evaluations.

Introduction: Preference Change and Preference Conflict

- Any theory of preference change will make an implicit or explicit reference to a decision-maker of the following sort:

$$A = \left\{ \begin{array}{cccc} A_0^0 & A_1^0 & \dots & A_k^0 \\ A_0^1 & A_1^1 & \dots & A_k^1 \\ \vdots & \vdots & A_j^i & \vdots \\ A_0^n & A_1^n & \dots & A_k^n \end{array} \right\}$$

where A_j^i is an agent represented by a set of preferences, k a time index and n an index of different decision situations, context or personalities.

Preference Change in the Multiple-Self

- Foundations for theories of preference change as well as explanations of preference change itself.
- Multiple-self models capture the stability of the decision-maker, both at a time and over time
- Clarifying notion of "self" for decision theory and preference change by offering a reductionist and non-reductionist definition of self

Agenda

1. The Simple Multiple-Self Model
2. The Extended Multiple-Self Model
3. Application: Intertemporal Weak Will

Multiple-self Models

- Different selves (parts, person-stages, agents) of a person as distinct, one person consists of a number of such selves.
- extreme end of psychological reductionism concerning personal identity over time

The Simple Multiple-Self Model

- A simple multiple-self model of personal identity over time is a tuple $M_s = \langle A, c \rangle$ where
  - $A = \{A_0, A_1 \ldots, A_k\}$ is a finite set of selves, drawn from some set Σ and k is a time index.
  - c is a connectedness function between pairs of selves $A_i, A_j \in A$, for instance $c : A \times A \rightarrow [0, 1]$

Interpreting Connectedness

- Interpretations from theories of personal identity over time
  - Psychological connectedness: similarity of psychological traits
  - Memory connectedness: similarity of memories
  - Sympathetic connectedness: stability of emotional caring
  - Physical connectedness: similarity of the body

Interpreting Connectedness

- Interpretations of connectedness and self:
  - Psychological reductionism: a self is a set of preferences, connectedness is degree of continuity of psychological features
  - Non-reductionism: a self is an agent described by a number of emotional, memory, psychological and physical features, connectedness is degree of continuity of such features

Psychological Connectedness

- Selves are sets of preferences, connectedness is determined by the proportion of unchanged preference relations
- Example:
  - Let A_1 have the preference ordering: $a \succ b \succ c \succ x \succ y \succ z$
  - Let A_4 have the preference ordering: $a \succ b \succ c \succ x \sim y \sim z$
  - The connectedness $c(A_1, A_4)$ can be determined by considering the distances between such orderings, for instance connectedness determined with Hamming distance: $c(A_1, A_4) = .8$

Applications of the Simple Multiple-Self Model

- Representation theorems for discounting factors via connectedness (Heilmann 2008): upon connectedness satisfying certain conditions it can be represented by a specific discounting function.
- Sufficient conditions for backward induction (Bach/Heilmann 2009): belief revision on the opponent's connectedness is used to deal with surprise information in reasoning in dynamic games

The Extended Multiple-Self

- An Extended Multiple-Self Model of personal identity is a tuple $M_e = \langle A, C_A \rangle$ where A is matrix of agents A^i_j:

$$A = \left\{ \begin{array}{cccc} A^0_0 & A^0_1 & \dots & A^0_k \\ A^1_0 & A^1_1 & \dots & A^1_k \\ \vdots & \vdots & A^i_j & \vdots \\ A^n_0 & A^n_1 & \dots & A^n_k \end{array} \right\}, k, n \in \mathbb{N}$$

and C_A a collection of connectedness measures between the agents such that

$$C_A = \left\{ c, d, \{c^0, c^1, \dots, c^n\}, \{d_0, d_1, \dots, d_k\} \right\}$$

Interpretations for the Extended Multiple-Self

- Again: reductionist and non-reductionist interpretations of connectedness
- Local connectedness functions are a measure of relative epistemic self-respect
- Can be used in deliberation about character planning according to the relative intertemporal credibility for utility maximization of internal viewpoints/evaluations

Intertemporal Weakness of the Will

- In intertemporal weakness of the will problems, a decision-maker is worse off by choosing a small short-term benefit over a larger long-term benefit.

Procrastination

	$\mathbf{t_1}$	$\mathbf{t_2}$
procrastinate	messing about	work to do
work diligently	working hard	being done

Drinks

	$\mathbf{t_1}$	$\mathbf{t_2}$
beer	get drunk	hangover
juice	sober	feeling healthy

Intertemporal Weak Will

	$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{\Sigma}$
- *procrastinate*	- messing about	- work to do	
$\star$ *beer*	$\star$ get drunk	$\star$ hangover	
	$\bigvee$	$\bigwedge$	$\bigwedge$
- *work diligently*	- working hard	- being done	
$\star$ *juice*	$\star$ sober	$\star$ feeling healthy	

Intertemporal Weak Will

	$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{\Sigma U_t}$
F	$U(F_1) = 3$	$U(F_2) = 1$	$\Sigma U(F_t) = 4$
G	$U(G_1) = 2$	$U(G_2) = 4$	$\Sigma U(G_t) = 6$

Intertemporal Weak Will

Definition (Weak and Strong Will)

Let F, G be acts and $U(F_1)$, $U(F_2)$, $U(G_1)$, $U(G_2)$ the utilities associated with these acts, respectively, at times 1 and 2. If

- $U(F_1) > U(G_1)$ and $U(G_2) > U(F_2)$ and
- $U(F_1) - U(G_1) \leq U(G_2) - U(F_2)$ then

an agent A has weak (strong) will if and only if $Choice_A = F$ ($Choice_A = G$).

Resolution of Intrapersonal Conflict

- use connectedness between selves in the multiple-self to characterize a decision-makers belief about his intertemporal stability
- Ainslie (1992, 2005) suggested the idea of understanding preference change as intrapersonal conflict between motivational states within a decision-maker
- Pettit (2004) proposed to "...[think] of the individual as a plurality of perspectives that interact in continuing search for the unity of a single, reasoned vision."

The Simple Multiple-Self Model

- Take the decision problem as outlined above. The decision-maker considers the two elements of each prospect separately and ascribes utilities to them.
- In a second step, the decision-maker considers the temporal dimension via simple psychological connectedness.

The Simple Multiple-Self Model

- $M_s = \langle A, c \rangle$, where $A = \{A_1, A_2\}$ and $c(A_1, A_2) = \{high, low\}$
  - let $c(A_1, A_2) = low$, for instance $c(A_1, A_2) = .25$.

F	$U(F_1) = 3$	$c(U(F_2)) = .25$	$\Sigma U(F_t) = 3.25$
G	$U(G_1) = 2$	$c(U(G_2)) = 1$	$\Sigma U(G_t) = 3$

  - Hence, F.
  - let $c(A_1, A_2) = high$, for instance $c(A_1, A_2) = .75$.

F	$U(F_1) = 3$	$c(U(F_2)) = .75$	$\Sigma U(F_t) = 3.75$
G	$U(G_1) = 2$	$c(U(G_2)) = 3$	$\Sigma U(G_t) = 5$

  - Hence, G.

The Simple Multiple-Self Model

- Low psychological connectedness can be a weight that diminishes the value of the options at t_2. Other types of non-reductionist interpretations of connectedness can also be used.
- If a decision-maker has evidence or otherwise reason to believe that his preferences will be unstable in the future, then this will diminish the present credibility of his present evaluations of future outcomes.
- This interpretation of weakness of will mimics the familiar time discounting rationalization of supposedly weak-willed choice.

The Simple Multiple-Self Model

Definition (Weak and Strong Will in the Simple Multiple-Self)

Let an agent A have a connectedness function $c[0,1]$ that measures the degree of preference change between times 1 and 2. Let F, G be acts and $U(F_1)$, $U(F_2)$, $U(G_1)$, $U(G_2)$ the utilities associated with these acts at times 1 and 2, respectively. If

- $U(F_1) > U(G_1)$ and $U(G_2) > U(F_2)$ and
- $U(F_1) - U(G_1) \leq c(U(G_2)) - c(U(F_2))$ then

an agent A has weak (strong) will if and only if $Choice_A = F$ ($Choice_A = G$).

The Extended Multiple-Self Model

- Consider a decision-maker that experiences a conflict as to whether he should choose F or G.
- In particular, he has two points of view: a *hedonic* one and a *prudent* one.
- Firstly, take the decision problem as outlined above. The decision-maker considers the two elements of each prospect separately and ascribed utilities to them.
- Secondly, the decision-maker considers the temporal dimension and two different viewpoints (hedonic and prudent) on this table via extended connectedness as follows.

The Extended Multiple-Self Model

- $M_e = \langle A, C_A \rangle$, where

$$A = \left\{ \begin{array}{cc} H_1 & H_2 \\ P_1 & P_2 \end{array} \right\}$$

  $C_A = \{c, c^*\}$ with $c(H_1, H_2) = \{high, low\}$ and $c^*(P_1, P_2) = \{high, low\}$

- if $c = c^* = low$ then F and if $c = c^* = high$ then G (when the intertemporal stability of the two viewpoints is the same, then the extended case collapses back into the simple case).

The Extended Multiple-Self Model

- if $c > c^*$ then F and if $c < c^*$ then G
- Let $c < c^*$, for instance $c(H_1, H_2) = .25$ and $c^*(P_1, P_2) = .75$

F	$U(F_1) = 3$	$c(U(F_2)) = .25$	$\Sigma U(F_t) = 3.25$
G	$U(G_1) = 2$	$c^*(U(G_2)) = 3$	$\Sigma U(G_t) = 5$

- Hence, G.
- Let $c > c^*$, for instance $c(H_1, H_2) = .75$ and $c^*(P_1, P_2) = .25$.

F	$U(F_1) = 3$	$c(U(F_2)) = .75$	$\Sigma U(F_t) = 3.75$
G	$U(G_1) = 2$	$c^*(U(G_2)) = 1$	$\Sigma U(G_t) = 3$

- Hence, F.

The Extended Multiple-Self Model

- When the intertemporal stability of the viewpoints differs, the evaluations corresponding to the choices that the viewpoints advocate, respectively, were weighted differently which leads to a higher intertemporal utility for one of the options (whichever is supported by the more stable viewpoint).

The Extended Multiple-Self Model

Definition (Weak and Strong Will in the Extended Multiple-Self)

Let an agent A have two connectedness functions $c[0,1]$ and $c^*[0,1]$ that measure the degree of preference change of her two personae $A_{Hedonic}$ and $A_{Prudent}$ between times 1 and 2. Let F, G be acts and $U(F_1)$, $U(F_2)$, $U(G_1)$, $U(G_2)$ the utilities associated with these acts at times 1 and 2, respectively. If

- $U(F_1) > U(G_1)$ and $U(G_2) > U(F_2)$ and
- $U(F_1) - U(G_1) \leq c^*(U(G_2)) - c(U(F_2))$ then

an agent A is governed by $A_{Hedonic}$ ($A_{Prudent}$) personality if and only if $Choice_A = F$ ($Choice_A = G$).

The Extended Multiple-Self Model

- Consider two initial evaluations that differ according to the two selves, hedonic and prudent.
- Prudent self:

	$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{\Sigma U_t}$
F	$U(F_1) = 3$	$U(F_2) = 1$	$\Sigma U(F_t) = 4$
G	$U(G_1) = 2$	$U(G_2) = 4$	$\Sigma U(G_t) = 6$

- Hedonic self:

	$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{\Sigma U_t}$
F	$U(F_1) = 4$	$U(F_2) = 2$	$\Sigma U(F_t) = 6$
G	$U(G_1) = 1$	$U(G_2) = 3$	$\Sigma U(G_t) = 4$

- Applying to each full table the respective local connectedness yields the same act as before.

Weak Will in the Multiple-Self

- Beliefs regarding the intertemporal stability of preferences can be used to characterise weakness of the will cases.
- In cases where there is evidence that the decision-maker is unstable as a whole (in the simple model) or where the hedonic preferences of the decision-maker are more stable (in the extended model), supposedly weak-willed acts can be rationalised.
- In contrast, if the decision-maker beliefs he really is stable (in the simple model) or the prudent preferences of the decision-maker are more stable (in the extended model), supposedly weak-willed acts are irrational.

Multiple-Self Models and Preference Change

- Multiple-selves are implicitly or explicitly assumed for any discussion of preference change or preference conflict
- Multiple-self models can be used in a reductionist way (self as a set of preferences) or non-reductionist way (self with more characteristics)
- Connectedness can be used for models of deliberation, as a decision-makers belief about the intertemporal stability of his agents.