Constructing the Formula of Universal Law

Matthew Braham and Martin van Hees

Amsterdam, 11 April 2012

Categorical Imperative 1: The Formula of Universal Law

Definition (The Formula of Universal Law)

Act only in accordance with that maxim through which you can at the same time will that it become a universal law.

CI1: The Formula of Universal Law

Definition (Step 1: Contradiction in Conception Test)

Can your maxim be a universal law?

- Perfect duties
- The examples of the *Grundlegung*:
  - False promises
  - Suicide

CI1: The Formula of Universal Law

Definition (Step 2: Contradiction in Will Test)

Given that your maxim can be a universal law, can you rationally *will* it to be so?

- Imperfect duties
- The examples of the *Grundlegung*:
  - Procrastination
  - Never helping others in need

Motivation

- Problems:
  - FUL does not generate definitive conclusions as to which actions have moral value and can be said to carry the force of obligation.
  - FUL does not yield the duties it is supposed to yield

Motivation

Conclusions about Kant's formula:

- 'radically defective' and 'pretty worthless' (Wood).
- 'a sad history of attempts ... no one has been able to make it work' (Herman).
- 'it may give either unacceptable guidance or none at all' (O'Neill).
- When used on its own, it cannot provide 'even a loose and partial action guide' (Hill).

Motivation

Workarounds:

1. The problem is one of interpretation – FUL has a *logical*, *teleological*, and *practical* interpretation (Korsgaard).
2. The informational structure of FUL needs to be specified (Rawls).
3. FUL needs to be augmented with anthropological assumptions about 'essential ends' (Korsgaard, Herman).
4. FUL needs to be reformulated (Parfit).

Motivation

- *A methodological observation:* None of these studies of FUL actually take up Kant's project on its own terms: to systematically examine its *formal* structure
- *Our project:* Propose a formal decision-theoretic framework for FUL and examine one part of it that is particularly contentious: the so-called 'Contradiction in Will Test' (CW-test)

The Framework

1. Game frames: worlds, games, actions, outcomes
   - W, N (cardinality n), games $G^w = (S_1 \ldots, S_n, R^n, \pi)$.
   - $\mathscr{D}^w \subseteq S_1 \times \ldots \times S_n$
2. Maxims
   - A mapping m_i that assigns to each world w an outcome-intention A_i^w and an action-intention T_i^w
3. Similarity of maxims
   - A reflexive and symmetric relation $\sim$ over the set of all individual maxims. Uniqueness.
   - A strategy of i *instantiates* a maxim of i in w if the strategy is an element of i's action-intention in that world. The combination of all strategies that instantiate a similar maxim m at w is $\mathscr{M}^w = T_1^w \times \ldots \times T_n^w$, where for all i, T_i^w is i's action-intention in w according to the maxim similar to m.

Rationality Requirements

Rationality requirements

- Intrapersonal consistency 1: Consistency between a person's maxims
- Intrapersonal consistency 2: Consistency of a person's maxim (proper fit between intended action and intended outcomes)
- Interpersonal consistency: CC and CW

The CC-Test

A maxim is conceptually inconsistent – fails the CC-test – if there is some world w such that not all of the individuals can act on the basis of that maxim in that world.

Definition (Contradiction in Conception (CC-test))

A maxim m of agent i contains a contradiction in conception (fails the CC-test) if and only if: $\mathscr{D}^w \cap \mathscr{M}^w = \emptyset$ for some world w.

The CW-Test

A maxim violates CW if universal adoption entails that the agent will not realize her intended outcome in some world ('practical contradiction').

Definition (Contradiction in the Will (CW-test))

A conceptually consistent maxim m of agent i contains a contradiction in the will (fails the CW-test) if and only if: for some w, and for all $s_N \in \mathscr{D}^w \cap \mathscr{M}^w$: $\pi(s_N) \notin A_i^w$.

Results 1

Definition (Complete Enforceability)

A maxim is completely enforceable if, and only if, for each state w the adoption of the action A^w ensures the realization of T^w.

Proposition

A maxim that is completely enforceable and conceptually consistent (i.e., passes the CC-test) never results in a contradiction in the will (i.e., always passes the CW-test). That is, any such maxim satisfies FUL.

Examples:

- Sidgwick's strong man
- The stoic retreat

Result 2

Definition (Agent-Neutral Maxims)

A maxim m_i of i is an *agent-neutral* maxim if, and only if, for all j and any m_j such that $m_i \sim m_j$: $A_i^w = A_j^w$ for all w.

Definition (Pure Consequentialism)

A maxim m of i is a *pure consequentialist* maxim if, and only if, for all w, $T_i^w = \{s \in S_i^w \mid \pi(s) \cap A_i^w \neq \emptyset\}$.

Proposition

Any maxim that is agent-neutral and purely consequentialist satisfies FUL.

Example: Utilitarianism

Results 3

Assume some solution concept Γ is given.

Definition (Sophisticated Consequentialism)

Given Γ, a maxim m of i is a sophisticated consequentialist maxim if, and only if, for any w,

(i) There is an equilibrium: T_i^w is set of all of i's eq actions at w and A_i^w is set of all eq outcomes at w;

(ii) There is no equilibrium: T_i^w is set of all of i's actions at w and A_i^w is set of all outcomes at w

Proposition

Any sophisticated consequentialist maxim satisfies FUL.

Example: Ethical egoism

Possible answers?

- Rendition of the CW-test is too weak
- The presumed counterexamples fail the CC-test
- Bite the bullet: accept that the CW-test doesn't do what it is supposed to do
- Chew on the bullet: "Comprehensive Kantianism"