Moral Norms in a Partly Compliant Society

Sebastian Kranz
University of Bonn

Paper in a nutshell

- Model interaction among selfish individuals & individuals with rational, limited moral motivation (only two types in basic model)
- Types are private information
- Model predicts wide range of stylized facts from economic experiments and is quite tractable.

Rational ethics in spirit of John C. Harsanyi

Rule Utilitarianism

Utilitarian Welfare Criterion	Welfare given by average expected utility of all individuals
&	
Rule-Consequentialism	Follow moral norms that maximize welfare if it becomes commonly known that morally motivated individuals follow these norms.

Rational ethics in spirit of John C. Harsanyi

Rule Utilitarianism

Utilitarian Welfare Criterion	Welfare given by average expected utility of all individuals
&	
Rule-Consequentialism	Follow moral norms that maximize welfare if it becomes commonly known that morally motivated individuals follow these norms.

We also consider an alternative welfare criterion

Complier Optimality	Welfare only puts explicit weight on expected utility of morally motivated types (called compliers).

Harsanyi's ideas have been largely ignored in ecnomic literature on social preferences

Related Literature

Rule Utilitarianism

- Harsanyi (e.g. 1977, 1985, 1992)
- Philosophical Literature on Rule-Utilitarianism (e.g. Brandt 1959)
- Feddersen and Sandroni (2006): A model of participation in elections

Social Preferences

- E.g. Levine (1998), Fehr & Schmidt (1999), Bolton & Ockenfels (2000), Charness & Rabin (2002), Dufwenberg & Kirchsteiger (2004), Falk & Fischbacher (2006), ...
- Lopez-Perez (2008): Norm-based approach, but no rule-consequentialistic norms.

Basic Model with two types (1)

- Underlying game with extensive form Γ
- n players
- Commonly known norm r specifies for each information set of Γ which actions are permitted or forbidden
- Selfish types: utility equal to payoff in underlying game
- Compliant types: feel disutility of $g > 0$ when playing at least once a forbidden action.
- Types are private knowledge
- Probability that a player is a complier denoted by κ (called complier's share) and common knowledge.

Basic Model (2)

- $\Gamma^*(r,\kappa,g,\Gamma)$ resulting game of imperfect information
- (σ^s,σ^c): strategy-profile of Γ^*

Definition: (σ^s,σ^c) is a *norm-equilibrium* for Γ,r,k,g if there exists beliefs μ such that $(\mu, (\sigma^s,\sigma^c))$ is a Perfect Bayesian equilibrium of $\Gamma^*(r,\kappa,g,\Gamma)$

- Multiple norm equilibria may exist for a given norm r. Let ψ be an equilibrium selection function.
- R set of norms for which norm equilibrium exists

Selfish types are never worse off than compliers

Proposition 1: In every norm equilibrium, selfish types have weakly higher expected utility than compliers:

$$U_i^c(.) \leq U_i^s(.) \forall\ i \text{ and } U^c(.) \leq U^s(.)$$

Intuition: Selfish types can always mimic compliant types

Rule-Utilitarian and Complier Optimal Norms

Complier optimal norm:

$$r^{o} \in \operatorname{argmax}_{r \in R}\{U^{c}(r,.)\}$$

Rule-utilitarian norm:

$$r^{utilitarian} \in \operatorname{argmax}_{r \in R}\{(1-\kappa)U^{s}(r,.)+\kappa U^{c}(r,.)\}$$

Rule-Utilitarian and Complier Optimal Norms

Complier optimal norm:

$$r^o \in argmax_{r \in R}\{U^c(r,.)\}$$

Rule-utilitarian norm:

$$r^{utilitarian} \in argmax_{r \in R}\{(1-\kappa)U^s(r,.)+\kappa U^c(r,.)\}$$

Generalization

- Consequentialistic welfare criterion: $W = W(r,\kappa,g,\Gamma,\psi)$
- W depends only on (expected) underlying payoffs, types, and (non-increasingly) on disutility from guilt.

Rule-consequentialistic norm for given W:

$$r^o \in argmax_{r \in R}\{W(r,.)\}$$

Rule-Consequentialistic Norm Equilibria and Compliance Principle

(σ^s,σ^c) is a *rule-consequentialistic norm equilibrium* (for given W) if it is selected under a rule-consequentialistic norm.

Similarly: rule-utilitarian and complier optimal norm equilibrium

Q: Any reason why a rule-consequentialistic norm should forbid an action that compliers play in the resulting norm equilibrium?

Rule-Consequentialistic Norm Equilibria and Compliance Principle

(σ^s,σ^c) is a *rule-consequentialistic norm equilibrium* (for given W) if it is selected under a rule-consequentialistic norm.

Similarly: rule-utilitarian and complier optimal norm equilibrium

Q: Any reason why a rule-consequentialistic norm should forbid an action that compliers play in the resulting norm equilibrium?

Compliance Principle: Let randomization devices be explicitedly included in underlying game Γ and equilibrium selection regular. For every rule-consequentialistic norm equilibrium ***(σ^s,σ^c)*** there is a rule-consequentialistic norm that (only) permits compliers' strategy-profile σ^c.

Example 1: Public Goods Game

- n players simultaneously choose contribution $c_i \in \mathbb{R}_0^+$
- Underlying payoffs: $u_i = \gamma \sum_{j=1}^{n} c_j - c_i$ with $1/n < \gamma < 1$

Example 1: Public Goods Game

- n players simultaneously choose contribution $c_i \in \mathbb{R}_0^+$
- Underlying payoffs: $u_i = \gamma \sum_{j=1}^{n} c_j - c_i$ with $1/n < \gamma < 1$

Rule-utilitarian norm: Every complier shall contribute

$$c^g \equiv g/(1-\gamma)$$

Compliers have negative expected payoff whenever $\kappa < \kappa_0 \equiv \frac{1-\gamma}{n\gamma-\gamma}$

Example 1: Public Goods Game

- n players simultaneously choose contribution $c_i \in \mathbb{R}_0^+$
- Underlying payoffs: $u_i = \gamma \sum_{j=1}^{n} c_j - c_i$ with $1/n < \gamma < 1$

Rule-utilitarian norm: Every complier shall contribute

$$c^g \equiv g/(1-\gamma)$$

Compliers have negative expected payoff whenever $\kappa < \kappa_0 \equiv \frac{1-\gamma}{n\gamma-\gamma}$

Complier optimal norm equilibrium: Contribute c^g if $\kappa \geq \kappa_0$ and 0 otherwise.

Example 2: Public Goods Game with Costly Punishment Technology

- Only 2 players
- 2nd stage: can reduce other players' payoff by $p_i \in \mathbb{R}_0^+$ if paying ϕp_i.
- Payoffs: $u_i = \gamma(c_1 + c_2) - c_i - \phi p_i - p_j$

Example 2: Public Goods Game with Costly Punishment Technology

- Only 2 players
- 2nd stage: can reduce other players' payoff by $p_i \in \mathbb{R}_0^+$ if paying ϕp_i.
- Payoffs: $u_i = \gamma(c_1 + c_2) - c_i - \phi p_i - p_j$

- Highest punishment that compliers would conduct: $p^g \equiv g/\phi$
- Selfish types contribute at most: $c^s := (\kappa p^g)/(1-\gamma)$

Proposition: The punishment technology increases contributions of every type by c^s in every rule-utilitarian and complier optimal norm equilibrium.

Example 2: Public Goods Game with Costly Punishment Technology

- Only 2 players
- 2nd stage: can reduce other players' payoff by $p_i \in \mathbb{R}_0^+$ if paying ϕp_i.
- Payoffs: $u_i = \gamma(c_1 + c_2) - c_i - \phi p_i - p_j$

- Highest punishment that compliers would conduct: $p^g \equiv g/\phi$
- Selfish types contribute at most: $c^s := (\kappa p^g)/(1-\gamma)$

Proposition: The punishment technology increases contributions of every type by c^s in every rule-utilitarian and complier optimal norm equilibrium.

- Assume players tremble with probability e and choose a random contribution c.
  - → All rule-utilitarian and complier optimal norms prescribe punishment that is proportional to the gap c^s-c.
- Furthermore: Intentions matter

Example 3: Sequential Prisoners' Dilemma

- 2 players sequentially contribute 1 or 0 units to a public good
- Assume $g \geq 1-\gamma$

Rule-utilitarian norm: Compliant player 1 always contributes. If $\kappa > (1-\gamma)/\gamma$ compliant player 2 conditionally cooperates and selfish player 1 contributes. If $\kappa < (1-\gamma)/\gamma$ compliant player 2 *unconditionally* contributes and selfish player 1 does not contribute.

Complier optimal norm: Compliant player 2 always conditionally cooperates. Selfish player 1 contributes if $\kappa > (1-\gamma)/\gamma$. A compliant player 1 contributes if $\kappa > \frac{1-\gamma}{3\gamma-1}$.

Example 4: Dictator Game (1)

- Player 1 can decide on split *(1-x,x)* between him and player 2.
- If both players have same linear utitility function in money, a zero-sum game...
  - → Every norm from which compliers do not deviate is a rule-utilitarian norm
  - → Complier optimality prescribes to act selfishly.
- No zero-sum game if players are risk- loss- or inequity averse.

Example 4: Dictator Game (2)

- Assume players are envious (loss averse with other player's payoff as reference level):

$$u_i(\pi) = \pi_i - \alpha \, max(\pi_j - \pi_i, 0)$$

Rule-utilitarian norm: Give *min(g,1 / 2)* whenever α > 0.

Complier optimal norm: Give *min(g,1 / 2)* whenever *α ≥ (1-κ) / 2κ*, otherwise give *0*.

Example 5: Ultimatum Game

- Substantial offers even if players are slightly inequity averse.
- Consider limit $\alpha \to 0$

Complier optimal norm equilibrium: Both compliers and selfish players contribute: $\min\{ 1/2, \kappa, g \}$.

- Intentions matter

Extensions

- Extended model with arbitrary number of types that differ in their degree of moral motivation
- A voting-by-feet model that gives additional justification for complier optimal norms (opposed to rule-utilitarian norms)

Summary: Model with rational, partly morally motivated types

Contribution to positive analysis:

- Tractable, robust model that is in line with many stylized facts across economic evidence.

Contribution to normative analysis:

- Complier optimality as alternative to rule-utilitarianism
- Describes how rational moral behavior concretly looks like in different situations under assumption that moral motivation differs
- Alternative implications for welfare analysis than models of social preferences

Extended Model with Multiple Types (1)

- Player's type $g \in G \subset R_0^+$
- A norm specifies permitted actions for each type: $r(h, g) \in A \mid h$
- Actions that are permitted for higher types must also be permitted for lower types.
- A type *g* feels disutility *g* if he plays at least once an action that is forbidden for his type.
- Norm equilibrium: $\{\sigma^g\}_{g \in G} \in \Sigma^G$

Results for Extended Model with Multiple Types

Rule-Utilitarian Norm in extended model:

Complier Optimal Norm in extended model:

Example: Public Goods Game:

Results for Extended Model with Multiple Types

A player's expected utility is weakly decreasing in her type

Modified version of compliance principle: There is always a rule-consequentialistic norm from which no type g > 0 deviates

Welfare under rule-consequentialistic norms generally increases if type distribution shifts towards higher levels of *g* (in other models of social preferences there is more ambiguity).