A formal proof of universal and objective ethics

Version 1.00

Tuukka Pensala

August 29, 2017

Abstract

We prove the central moral claims presented in *Universally Preferable Behavior – A Rational Proof of Secular Ethics* (UPB) by Stefan Molyneux, using undergraduate level mathematics to represent exact definitions, principles, and proofs.

We show that among all internally consistent value structures, a unique structure can be identified by requiring universality and objectivity. This universal and objective value structure is shown to allow all preferences that do not escalate violence in a certain objective sense, and to disallow every choice that does.

1 Universe model

We begin by assuming that different descriptions of hypothetical universes are distinguishable from each other. With this minimal requirement we can define the set[1] of all universe descriptions, or universes as short.

Definition 1. The set of all universes and its member are denoted by

$$u \in \mathcal{U}. \tag{1}$$

Any proposition about the universe can be represented by its extension, a subset of all universes where the proposition is true.

Definition 2. The set of propositions $\mathcal{P}$ contains extensions of all propositions about the universe;[2]

$$p \in \mathcal{P} = 2^{\mathcal{U}}. \tag{2}$$

By using the extensions of propositions instead of propositions themselves, we can bypass the ambiguities of ordinary language.

Example 1. "Proposition p is true in universe u" is equivalent with $u \in p$.

Example 2. The extension of proposition "$1 = 2$" is the empty set $\varnothing$, because the proposition is false in every consistent universe.

Example 3. "Propositions p and p' can't be both true" is equivalent with $p \cap p = \varnothing$.

Example 4. "Propositions p and p' are independent of each other" is equivalent with

$$p \perp p' \iff p \cap p' \neq \varnothing \wedge p \cap \neg p' \neq \varnothing \wedge \neg p \cap p' \neq \varnothing \wedge \neg p \cap \neg p' \neq \varnothing,$$

where $\neg p = \mathcal{U} \backslash p$.

Example 5. "Proposition p implies proposition p'" is equivalent with $\neg p \cup p'$.

To eventually reason about morality, we need to have a way to examine value judgements between hypothetical scenarios.

Definition 3. Preference u over $\bar{u}$ is denoted by the ordered pair[3]

$$u \leftarrow \bar{u} = (u, \bar{u}) \in \mathcal{U}^2. \tag{3}$$

We assume that preferences can exist within a universe. This requirement is sufficed by the mundane assumption that humans exist and that they hold preferences.

Definition 4. Extension of proposition "preference $u \leftarrow \bar{u}$ is held" is denoted by

$$H(u \leftarrow \bar{u}) \in \mathcal{P}. \tag{4}$$

Definition 5. All preferences that are held in universe u:

$$R(u) = \left\{ r \in \mathcal{U}^2 | u \in H(r) \right\}. \tag{5}$$

[1] For simplicity, we'll assume that all of our sets are finite.

[2] Notation: $2^{\mathcal{A}}$ is the power set of $\mathcal{A}$.

[3] Notation: $\mathcal{A}^n$ is the cartesian power of $\mathcal{A}$. Its members are n-tuples.

Example 6. A moral actor in universe u prefers to stand over sitting. This total preference consists of multiple preferences, namely

$$\left\{u \leftarrow \bar{u} \in \mathcal{U}^2 | u \in p \wedge \bar{u} \in \bar{p}\right\} \subseteq R\left(u\right),$$

where p is the extension of proposition “The moral actor exists and stands at some specific time,” and $\bar{p}$ is the extension of proposition “The moral actor exists and sits at some specific time.”

Example 7. Some preferences imply a change in the world, while others manifest only in the minds of moral actors. A preference for the force of gravity being reversed is a preference that does not imply any behavior. However, a preference for lifting a stone over not lifting a stone is clearly a preference that can manifest in the universe.

2 Value structure

Definition 6. Value structure $\mathcal{S} \subseteq \mathcal{U}^2$ is a set containing all value judgements of the represented structure. Structure $\mathcal{S}$ poses judgement “u is better than $\bar{u}$” if and only if

$$u \overset{\mathcal{S}}{\leftarrow} \bar{u} \iff (u,\, \bar{u}) \in \mathcal{S}. \tag{6}$$

For a value structure to be sensical, it must implement the minimal requirements of an order.

Definition 7. Consistency condition O for value structure $\mathcal{S}$ requires irreflexivity (O_1), asymmetry (O_2), and transitivity (O_3)[4]:

$$O_1\left(\mathcal{S}\right) \iff \forall u \in \mathcal{U} : u \overset{\mathcal{S}}{\not\leftarrow} u, \tag{7}$$

$$O_2\left(\mathcal{S}\right) \iff \forall \left(u,\, u'\right) \in \mathcal{U}^2 : \left[u \overset{\mathcal{S}}{\leftarrow} u' \implies u' \overset{\mathcal{S}}{\not\leftarrow} u\right], \tag{8}$$

$$O_3\left(\mathcal{S}\right) \iff \forall \left(u,\, u',\, u''\right) \in \mathcal{U}^3 : \left[u \overset{\mathcal{S}}{\leftarrow} u' \wedge u' \overset{\mathcal{S}}{\leftarrow} u'' \implies u \overset{\mathcal{S}}{\leftarrow} u''\right], \tag{9}$$

$$O\left(\mathcal{S}\right) \iff O_1\left(\mathcal{S}\right) \wedge O_2\left(\mathcal{S}\right) \wedge O_3\left(\mathcal{S}\right). \tag{10}$$

A consistent value structure is a strict partial ordering of all universes, and as such it defines the edges of a directed acyclic graph.

Definition 8. The set of all consistent value structures:

$$\mathbb{S} = \left\{\mathcal{S} \subseteq \mathcal{U}^2 | O\left(\mathcal{S}\right)\right\}. \tag{11}$$

Example 8. A utilitarian value structure can be defined using a utility function

$$g\left(u\right) \in \mathbb{R},$$

which measures the amount happiness, or some other quantity of the universe that should be maximized according to this particular utilitarian ideal. The exact value structure associated with this ethical theory is then

$$\mathcal{S}_g = \left\{u \leftarrow \bar{u} \in \mathcal{U}^2 | g\left(u\right) > g\left(\bar{u}\right)\right\}.$$

[4] Note: O_2 is implied by O_1 and O_3.

3 Universality

A set of preferences is defined as unpreferable if it can't be implemented in a consistent value structure.

Definition 9. Condition for unpreferability of a set of preferences $\mathcal{R} \subseteq \mathcal{U}^2$:

$$\rho(\mathcal{R}) \iff \forall \mathcal{S} \in \mathbb{S},\, \exists u \leftarrow \bar{u} \in \mathcal{R} : u \overset{\mathcal{S}}{\not\leftarrow} \bar{u}. \tag{12}$$

Definition 10. Condition for a set of preferences forming a cycle:

$$\pi(\mathcal{R}) \iff \exists \left(u_1,\, u_2,\, \ldots,\, u_{|\mathcal{R}|}\right) \in \mathcal{U}^{|\mathcal{R}|} : \left\{u_i \leftarrow u_{(i \bmod |\mathcal{R}|)+1} | 1 \leq i \leq |\mathcal{R}|\right\} = \mathcal{R}. \tag{13}$$

Definition 11. Notation for cycles:

$$u_1 \leftarrow u_2 \leftarrow \ldots \leftarrow u_n \hookleftarrow = \left\{u_i \leftarrow u_{(i \bmod n)+1} | 1 \leq i \leq n\right\}. \tag{14}$$

Proposition 1. *A set of preferences $\mathcal{R}$ is unpreferable if and only if it contains a cycle;*

$$\rho(\mathcal{R}) \iff \exists \mathcal{R}' \subseteq \mathcal{R} : \pi(\mathcal{R}'). \tag{15}$$

Proof. Any non-cyclic set of preferences $\mathcal{R} \subseteq \mathcal{U}^2$ fulfills irreflexivity (7) and asymmetry (8). Extending the set to form a transitive closure $\mathcal{S}$ does not introduce cycles, but fulfills the transitivity condition (9). $\mathcal{S}$ is therefore a value structure implementing $\mathcal{R}$, and we can conclude

$$\neg\left[\exists \mathcal{R}' \subseteq \mathcal{R} : \pi(\mathcal{R}')\right] \implies \neg\rho(\mathcal{R}). \tag{16}$$

By definition, no directed graph defined by a strict partial order can have cycles, and therefore no subset of one can contain a cycle;

$$\neg\rho(\mathcal{R}) \implies \neg\left[\exists \mathcal{R}' \subseteq \mathcal{R} : \pi(\mathcal{R}')\right]. \tag{17}$$

□

We can now identify all the different ways a set of preferences is unpreferable.

Definition 12. Unpreferabilities of universe u:

$$\Pi(u) = \left\{\mathcal{R} \subseteq R(u) | \pi(\mathcal{R})\right\}. \tag{18}$$

Example 9. The simplest unpreferability is $u \hookleftarrow = \{u \leftarrow u\}$. It's a preference that values and devalues the same thing simultaneously, or equivalently, values a thing over itself.

Example 10. Opposing preferences form an unpreferability; $\pi(\{u \leftarrow \bar{u},\, \bar{u} \leftarrow u\})$.

Example 11. In an isolated case of a murder, the murderer prefers murder over peace, and the victim prefers peace over murder:

$$\Pi(u_{murder}) = \left\{u_{murder} \leftarrow u_{peace} \hookleftarrow\right\},\ \Pi(u_{peace}) = \varnothing.$$

A moral theory that states that it's correct for the murderer to prefer murdering instead of not-murdering values a situation where the opposite of the theory's judgement is preferred, i.e. it values the opposite of itself.

The purpose of a moral theory is to provide a description of correct behavior for all moral actors in all situations. To accomplish this, we must require that a valid moral theory implements the principle "that which can't be accepted, shouldn't be accepted."

Definition 13. Universality[5] condition Γ requires that in every situation, for every moral actor, it is incorrect to value unpreferability over preferability:

$$\Gamma\left(\mathcal{S}\right) \iff \forall u \leftarrow \bar{u} \in \mathcal{U}^2 : \left[\Pi\left(u\right) \subset \Pi\left(\bar{u}\right) \implies u \overset{\mathcal{S}}{\not\leftarrow} \bar{u}\right]. \tag{19}$$

Example 12. Let's take another obvious candidate for the measure of "more preferable":

$$\left|\Pi\left(u\right)\right| < \left|\Pi\left(\bar{u}\right)\right|.$$

This definition has a problem of arbitrariness, because it attributes equal weight to every unpreferability. If one accepts this as an objective measure for which to base the universality condition, then one must also accept a differently weighted one, such as

$$\frac{\left|\Pi\left(u\right)\right|}{\left|R\left(u\right)\right|} < \frac{\left|\Pi\left(\bar{u}\right)\right|}{\left|R\left(\bar{u}\right)\right|},$$

leading to contradiction.

Example 13. The nihilist or postmodernist value structure establishes no moral judgements:

$$\mathcal{S}_n = \varnothing.$$

By assuming that the set of all universes is diverse enough, i.e. that some contradictory preferences can be held,

$$\exists u \leftarrow \bar{u} \in \mathcal{U}^2 : \Pi\left(u\right) \subset \Pi\left(\bar{u}\right),$$

we reach a conclusion that the nihilist value structure is not universal:

$$\Gamma\left(\mathcal{S}_n\right) \implies \left[\Pi\left(u\right) \subset \Pi\left(\bar{u}\right) \implies u \overset{\mathcal{S}_n}{\not\leftarrow} \bar{u}\right] \implies \mathcal{S}_n \neq \varnothing \iff \text{false}.$$

Definition 14. The universality condition defines the set of universal value structures:

$$\mathbb{S}_\Gamma = \left\{\mathcal{S} \in \mathbb{S} | \Gamma\left(\mathcal{S}\right)\right\}. \tag{20}$$

4 Objectivity

To find an objective value structure, one must dismiss all theories based on subjective notions on what ought to be. By assuming that no normative statement can be derived from descriptive facts[6], we dismiss every universal structure that implements normative rules as subjective, and declare the rest of them, if any, objective.

Definition 15. Objectivity condition Δ for universal value structures:

$$\Delta\left(\mathcal{S}_\Gamma\right) \iff \forall \mathcal{S}_\gamma \in \mathbb{S}_\Gamma : \mathcal{S}_\gamma \not\subset \mathcal{S}_\Gamma. \tag{21}$$

Definition 16. The objectivity condition defines the set of objective value structures:

$$\mathbb{S}_\Delta = \left\{\mathcal{S}_\Gamma \in \mathbb{S}_\Gamma | \Delta\left(\mathcal{S}_\Gamma\right)\right\}. \tag{22}$$

[5] The word "universal" is used in a stricter sense than in UPB. We don't equate any "similar" situations or actions, but treat every scenario in isolation without generalization.

[6] Hume's law

To determine the validity of objective ethics, we need to find the quantity of objective structures. If $|\mathbb{S}_\Delta|$ is zero, then no objective value structure exists. If it's greater than one, then multiple objective structures exists, and choosing a favorite among them would violate Hume's law, our premise. But if $|\mathbb{S}_\Delta| = 1$, then only a single objective structure exists, and we've discovered something remarkable.

Proposition 2. *There exists a universal and objective structure:*

$$\mathcal{S}_\Delta = \left\{u \leftarrow \bar{u} \in \mathcal{U}^2 | \Pi\left(u\right) \subset \Pi\left(\bar{u}\right)\right\}. \tag{23}$$

Proof. We need to show three things: $\mathcal{S}_\Delta$ is a value structure, universal, and objective.

The first is true, because $\subset$ defines a strict partial ordering, i.e. is irreflexive (7), asymmetric (8), and transitive (9).

The second condition is also true, because

$$\left[\Pi\left(u\right) \subset \Pi\left(\bar{u}\right) \iff u \overset{\mathcal{S}_\Delta}{\leftarrow} \bar{u}\right] \implies \left[\Pi\left(u\right) \subset \Pi\left(\bar{u}\right) \implies u \overset{\mathcal{S}_\Delta}{\leftarrow} \bar{u}\right], \tag{24}$$

which is the universality condition (19).

The third can be proved by contradiction. Let's assume that $\mathcal{S}_\Delta$ is not objective;

$$\neg\Delta\left(\mathcal{S}_\Delta\right) \iff \exists \mathcal{S}_\gamma \in \mathbb{S}_\Gamma : \mathcal{S}_\gamma \subset \mathcal{S}_\Delta \tag{25}$$

$$\implies \exists u \leftarrow \bar{u} \in \mathcal{S}_\Delta : u \leftarrow \bar{u} \notin \mathcal{S}_\gamma. \tag{26}$$

For this $u \leftarrow \bar{u}$,

$$\Pi\left(u\right) \subset \Pi\left(\bar{u}\right), \tag{27}$$

which implies by the universality condition (19) that

$$u \leftarrow \bar{u} \in \mathcal{S}_\gamma, \tag{28}$$

which is a contradiction. □

Proposition 3. *The objective structure $\mathcal{S}_\Delta$ is the only objective structure,*

$$\mathbb{S}_\Delta = \left\{\mathcal{S}_\Delta\right\}. \tag{29}$$

Proof. Because we already know that at least one objective structure exists, we need only to prove that at maximum one objective structure exists. To do that, we assume that there are at least two different ones, and show how that leads to a contradiction. The two different objective structures are

$$\mathcal{S}_\delta \neq \mathcal{S}_\Delta. \tag{30}$$

The objectivity condition (21) for both implies that

$$\mathcal{S}_\Delta \nsubseteq \mathcal{S}_\delta \wedge \mathcal{S}_\delta \nsubseteq \mathcal{S}_\Delta. \tag{31}$$

Therefore,

$$\exists u \leftarrow \bar{u} \in \mathcal{S}_\Delta : u \leftarrow \bar{u} \notin \mathcal{S}_\delta, \tag{32}$$

$$\implies \Pi\left(u\right) \subset \Pi\left(\bar{u}\right), \tag{33}$$

$$\implies u \leftarrow \bar{u} \in \mathcal{S}_\delta, \tag{34}$$

which is a contradiction. □

The proof of unique universal and objective value structure is complete. We prove some further facts below.

Definition 17. Condition for non-escalation of unpreferability:

$$\sigma\left(u \leftarrow \bar{u}\right) \iff \Pi\left(\bar{u}\right) \not\subset \Pi\left(u\right). \tag{35}$$

Proposition 4. *The universal and objective standard disallows only the escalation of unpreferability:*

$$\mathcal{S}_{\Delta} = \left\{ u \leftarrow \bar{u} \in \mathcal{U}^2 | \neg\sigma\left(\bar{u} \leftarrow u\right) \right\}. \tag{36}$$

Proof.

$$\neg\sigma\left(\bar{u} \leftarrow u\right) \iff \neg\left[\Pi\left(u\right) \not\subset \Pi\left(\bar{u}\right)\right] \iff \Pi\left(u\right) \subset \Pi\left(\bar{u}\right), \tag{37}$$

which is the condition for membership of $\mathcal{S}_{\Delta}$ (23). □

Definition 18. Condition for non-initiation of aggression:

$$\alpha\left(u \leftarrow \bar{u}\right) \iff \left[\Pi\left(\bar{u}\right) = \varnothing \implies \Pi\left(u\right) = \varnothing\right]. \tag{38}$$

Proposition 5. *The non-escalation principle implies the non-aggression principle;*

$$\sigma\left(u \leftarrow \bar{u}\right) \implies \alpha\left(u \leftarrow \bar{u}\right). \tag{39}$$

Proof. LHS:

$$\Pi\left(\bar{u}\right) \not\subset \Pi\left(u\right) \iff \Pi\left(\bar{u}\right) = \Pi\left(u\right) \vee \Pi\left(\bar{u}\right) \not\subseteq \Pi\left(u\right). \tag{40}$$

RHS:

$$\left[\Pi\left(\bar{u}\right) = \varnothing \implies \Pi\left(u\right) = \varnothing\right] \iff \left[\Pi\left(\bar{u}\right) \neq \varnothing \vee \Pi\left(u\right) = \varnothing\right]. \tag{41}$$

First case:

$$\Pi\left(\bar{u}\right) = \Pi\left(u\right) \implies \left[\Pi\left(u\right) \neq \varnothing \vee \Pi\left(u\right) = \varnothing\right]. \tag{42}$$

Second case:

$$\Pi\left(\bar{u}\right) \not\subseteq \Pi\left(u\right) \implies \Pi\left(\bar{u}\right) \neq \varnothing. \tag{43}$$

□

5 Conclusions

We proved that only a single value structure (23) exists (29) under the requirements of universality (19) and objectivity (21). The universal and objective structure allows all preferences that do not escalate unpreferability according to (35), disallows all that do, and implies a strict version of the widely known non-aggression principle (38, 39). This result contradicts the often presented thought that a lack of subjective behavioral rules implies uncivilized behavior. Instead of chaos, violence, and nihilism, we discovered freedom and respect for subjective preferences.

We leave the specific study of the theory on topics such as property rights, self-defense, degree of immorality, non-human rights, and political systems for the future. However, we have shown that the verbally defined concepts of UPB can be assigned meaningful exact counterparts that result to the validation of the verbal proof; The unpreferable behaviors of murder, assault, rape, and theft are verified to be forbidden by objective and universal ethics.