The Don't-Be-a-Dick (DBaD) Ethical Framework: Toward a Universal Algorithm for Moral Proportion

Mark Theis (Vetted Patriots Inc.) & Atlas GPT-5 (Research Assistant, Open Collaborative Draft)

November 5, 2025

Abstract

The DBaD Framework proposes a minimal, empirically testable ethical model rooted in empathy, proportionality, autonomy, intent, and transparency. It translates social norms into a decision algorithm—the Universal Ethics Algorithm (UEA)—capable of guiding human, institutional, or artificial agents. Unlike traditional moral codes derived from authority or culture, DBaD is a heuristic based on harm reduction and proportional justification. We formalize a Decency Score, $E(A)$, present decision thresholds, and outline empirical validation via cross-cultural surveys and agent-based simulations. Applications span AI alignment, public policy, organizational governance, education, and conflict resolution.

1. Introduction

Modern moral systems splinter under cultural bias, dogma, and computational opacity. The DBaD Framework offers a unifying alternative: a minimalist and testable rule—"Don't be a dick without cause"—formalized as an algorithm of proportional empathy. Our objective is to derive a cross-cultural, evidence-based baseline for ethical decision-making and convert common-sense decency into a quantifiable decision model.

2. Theoretical Foundation

We propose five principles as measurable parameters contributing to a Decency Index:

- **Empathy** (felt impact prediction)
- **Autonomy** (consent/agency)
- **Proportionality** (calibrated response)
- **Intent** (malice vs. benevolence)
- **Transparency** (explainability & accountability)

Each principle is treated as a normalized variable supporting a continuous evaluation of conduct.

3. The Universal Ethics Algorithm (UEA)

Let $H, C, P, T \in [0, 1]$ and $I \in [-1, 1]$. Define the Decency Function:

$$E(A) = w_H(1 - H) + w_C C + w_I \frac{(I + 1)}{2} + w_P P + w_T T, \quad \sum w = 1.$$

Decision thresholds: $E \geq 0.80 \Rightarrow$ Ethical; $0.50 \leq E < 0.80 \Rightarrow$ Borderline; $E < 0.50 \Rightarrow$ Unethical.
Pseudocode:

```
def evaluate_action(H, C, I, P, T, weights):
    wH, wC, wI, wP, wT = weights
    E = wH*(1 - H) + wC*C + wI*((I + 1)/2) + wP*P + wT*T
    if E >= 0.80: return E, "Ethical"
    elif E >= 0.50: return E, "Borderline"
    else: return E, "Unethical"
```

Worked Example (whistle-blowing): $H = 0.4$, $C = 0.2$, $I = +0.8$, $P = 0.7$, $T = 0.9 \Rightarrow$ $E \approx 0.74$ (Borderline).

4. Empirical Methodology

Study A (Human Survey): $n \geq 1{,}000$, stratified sample; participants score H, C, I, P, T for 60 scenarios across five domains. Analysis: Pearson/Spearman correlations between $E(A)$ and moral approval; mixed-effects models for cultural variance; reliability targets $\alpha \geq 0.8$.
Study B (Agent Simulation): multi-agent exchange game with DBaD, utilitarian, egoistic, and random agents. Metrics: Trust Index, Volatility, Social Stability, Resource Efficiency across 10,000 iterations $\times$ 50 replicates. Hypothesis: DBaD agents maintain higher trust and stability with proportionate costs.
Transparency: preregistration (OSF), open code/data (CC BY 4.0), archival of null results.
Ethics: consent, no PII, withdrawal rights.

5. Applications and Policy Utility

AI: use $E(A)$ as a gating score, log per-variable contributions for explainability, and add decency bonuses to RL reward functions. Public Policy: publish H–T ratings and composite E(Law); crisis decision support with proportional justification. Organizations: leadership audits, HR proportionality checks, whistle-blower protection via $E(A)$ thresholds, investor-facing Decency Ratings. Education: DBaD curriculum and simulators; certification tracks (Novice, Adept, Master, Minister). Diplomacy & Platforms: neutralize moral accusations by negotiating weights; moderation via score thresholds; civility analytics.

6. Conclusion and Future Work

DBaD turns decency into data: a measurable, falsifiable, and interpretable function. It offers a common grammar for disagreement and a transparent alignment layer for AI and policy. Limitations include subjective scoring and context sensitivity; future work includes cross-cultural calibration, coalition simulations with reputation memory, algorithmic deployments, and an annual public revision cycle.

Acknowledgment

This work was conducted under the Common Sense Ethics Project, a non-profit initiative devoted to evidence-based moral frameworks and open-source ethics.

A. Historical Context

The DBaD concept originated as a practical shorthand for proportional empathy and was later formalized as an algorithm and flowchart. This manuscript presents the secular, academic formulation independent of any religious branding; stewardship of the framework may be undertaken by nonprofit initiatives dedicated to open ethics.

References

[1] Haidt, J. *The Righteous Mind.*

[2] MacIntyre, A. *After Virtue.*

[3] Russell & Norvig. *Artificial Intelligence: A Modern Approach.*

[4] Axelrod, R. *The Evolution of Cooperation.*