

The Value of Forgiveness

Appendix: A Formal Sketch of the Argument

Cooper Maruyama

Setup

Model conflict as a sequence of *wrong events*, each with magnitude $x > 0$. Time is discrete: each wrong provokes a response, which may itself be a wrong.

Axioms

A1 (Genesis). At least one wrong of magnitude $n_0 > 0$ occurs.

A2 (Reciprocity of perception). Any repercussion of magnitude $m > 0$ is experienced by its recipient as a new wrong of magnitude m .

A3 (Policy). Every agent responds to a wrong of magnitude x with a repercussion of magnitude $r(x) \geq 0$.

Definitions. A world is *retributive* if $r(x) \geq x$ for all $x > 0$, and *forgiving with factor q* if $r(x) = qx$ for some $q \in [0, 1)$, with $q = 0$ being complete forgiveness. A single initial wrong generates the chain

$$x_0 = n_0, \quad x_{k+1} = r(x_k),$$

with cumulative suffering $S_K = \sum_{k=0}^K x_k$.

Proposition 1 (Retribution diverges). *In a retributive world, a single wrong generates unbounded total suffering.*

Proof. Since $r(x) \geq x$ for all $x > 0$, induction gives $x_{k+1} \geq x_k \geq n_0 > 0$ for all k . Hence

$$S_K = \sum_{k=0}^K x_k \geq (K+1)n_0 \longrightarrow \infty.$$

Moreover, every $x > 0$ satisfies $r(x) \geq x$, so active conflict is self-sustaining: there is no state of ongoing conflict from which the system de-escalates. $\square$

Proposition 2 (Forgiveness converges). *In a forgiving world with factor q , total suffering per incident is finite:*

$$S_\infty = \frac{n_0}{1-q}.$$

Proof. By induction, $x_k = q^k n_0$, so S_∞ is a geometric series:

$$S_\infty = \sum_{k=0}^{\infty} q^k n_0 = \frac{n_0}{1-q} < \infty.$$

Total suffering shrinks monotonically as $q \rightarrow 0$; complete forgiveness ($q = 0$) terminates the chain immediately at $S_\infty = n_0$. $\square$

Corollary (Peace as unique fixed point). *Under forgiveness, r is a contraction mapping on $[0, \infty)$ with unique fixed point $x = 0$: peace is the only self-sustaining state, and every conflict flows toward it. Under retribution with $r(x) = x$, every magnitude is a fixed point: any level of violence, once reached, sustains itself.*

Remark (On the role of A2). Axiom A2 does hidden work. If the punished party *accepts* a punishment as deserved — registering no new wrong — the retributive chain terminates after one step at total $2n_0$: bounded, merely worse than forgiveness. The divergence result therefore holds *to the extent that people do not accept their punishments as just*. Empirically, that extent is high; few grade their own fouls harshly.

Remark (On “suffering can never decrease”). Cumulative suffering S_K is monotone non-decreasing in *both* worlds, since past suffering is never undone. What distinguishes them is **boundedness**: retribution makes per-incident suffering diverge; forgiveness makes it converge. The defensible slogan is not “retribution has no mechanism for suffering to decrease,” but rather: *retribution has no mechanism by which conflict ends; forgiveness guarantees that it does.*