

THE RECIPROCAL SACRIFICE TEST

Source work: library/sacrifice.pdf

THE RECIPROCAL SACRIFICE TEST

THE RECIPROCAL

SACRIFICE TEST

AND THE RECIPROCAL PRIORITY VERIFICATION PROTOCOL

A method for distinguishing what an agent professes, reflectively endorses, actually follows, and morally ought to follow

MICHAEL RICHARD HAIMES

What one accepts sacrificing is ranked lower; what one refuses to sacrifice governs.

THE RECIPROCAL SACRIFICE TEST

Abstract

When an agent sincerely holds two commitments that cannot both be honored, ordinary statements of value may fail to reveal which commitment is meant to govern. Actual behavior is stronger evidence of what prevailed, but behavior can be ambiguous because it may reflect coercion, restricted options, habit, incomplete information, weakness of will, or an unacknowledged third consideration. This paper presents the Reciprocal Sacrifice Test, a controlled procedure for identifying reflectively endorsed priority between two conflicting commitments. The test constructs two matched alternatives: one in which commitment A is preserved at the accepted cost of frustrating commitment B, and a reciprocal alternative in which B is preserved at

the accepted cost of frustrating A. The comparison is repeated with neutral wording and reversed presentation order, while relevant differences in severity, probability, duration, reversibility, responsibility, consent, and effects on third parties are controlled as far as possible. Graduated changes in cost locate the threshold at which the agent's priority reverses. The test therefore permits findings of strict priority, conditional priority, approximate equality, parity, incomparability, instability, or indeterminacy. The paper then combines the test with the core insight of revealed preference theory, introduced by economist Paul A. Samuelson in 1938 and developed under that name in 1948. Reciprocal

reflection identifies what an agent endorses as governing; observed choice supplies evidence of what actually governs. Their combination, followed by discrepancy analysis and independent moral evaluation, forms the Reciprocal Priority Verification Protocol. The protocol distinguishes professed priority, reflectively endorsed priority, behaviorally effective priority, and morally justified priority. It is proposed as a strongest-in-class conceptual pathway for the narrow task of diagnosing two-commitment conflicts across reflection, action, and justification, while remaining subject to future empirical validation.

CENTRAL CLAIM

When two commitments cannot both be honored, their reflectively endorsed priority is revealed by testing which commitment the agent will preserve at the accepted cost of frustrating the other - and then reversing the sacrifice under matched conditions.

- 1. The philosophical problem

Two commitments may both be sincere and logically compatible in ordinary circumstances yet become practically incompatible in a particular case. The conflict is not necessarily between two propositions that cannot both be true. It is between two commitments that cannot both govern the same decision. Example: A person may sincerely say both, "I want you to be happy" and "I do not want your conduct to make me angry." The philosophical question arises only when the relevant happiness and the avoidance of anger cannot both be preserved.

THE RECIPROCAL SACRIFICE TEST

The Reciprocal Sacrifice Test is not primarily a test of factual truth. It is a test of practical priority: which commitment the agent reflectively authorizes to survive when one must bear the cost of the other.

- 2. The Reciprocal Sacrifice Test

Definition

THE TEST

The Reciprocal Sacrifice Test identifies the reflectively endorsed priority between two conflicting commitments by testing each commitment both as the value preserved and as the value sacrificed under sufficiently matched conditions.

Procedure

- 1. Identify commitments A and B independently, charitably, and at the same level of generality.
- 2. Establish that a genuine conflict exists. If a reasonable option preserves both commitments, priority

need not yet be forced.

- 3. Construct outcome A: preserve A while accepting the frustration of B.
- 4. Construct outcome B: preserve B while accepting the frustration of A.
- 5. Match relevant conditions, including severity, duration, probability, reversibility, responsibility,

consent, action versus omission, and consequences for third parties.

- 6. Present both alternatives neutrally, reverse their order, and rephrase them to test resistance to

framing.

- 7. Vary the magnitude of each cost until the preference reverses or reaches a stable boundary. This

identifies a priority threshold.

- 8. Record the selected commitment, the reason, the level of confidence, the threshold, and any

circumstances that would change the result. Formal statement Let A and B be commitments held by person P in context C. Let O(A) be the outcome in which A is preserved and B is frustrated. Let O(B) be the outcome in which B is preserved and A is frustrated.

FORMAL RESULT

If P stably and reflectively selects O(A) over O(B) under sufficiently matched conditions, then A has greater reflectively endorsed priority than B for P in context C. The qualification "in context C" is essential. Human commitments are often conditional. A person may accept mild anger for another person's substantial happiness but refuse severe and lasting harm for a minor pleasure. The result is then a threshold-sensitive priority rather than an unconditional hierarchy.

THE RECIPROCAL SACRIFICE TEST

Possible findings

Finding

Meaning

A has strict priority A is preserved at B's cost, but the reverse sacrifice is rejected. B has strict priority B is preserved at A's cost, but the reverse sacrifice is rejected. Conditional priority The ordering changes when the cost crosses a threshold. Approximate equality The reciprocal sacrifices are treated as practically equivalent.

Parity

The commitments are comparable, but neither is better, worse, or exactly equal.

Incomparability

No justified common basis of comparison is available.

Instability

Minor changes in wording, order, or mood reverse the result.

Indeterminacy

The evidence is insufficient to establish a ranking.

- 3. Worked example: happiness and anger

Commitment A: "I want you to experience substantial and legitimate happiness." Commitment B: "I want to avoid becoming angry because of your conduct."

RECIPROCAL QUESTION A

Would I accept becoming angry if that were the unavoidable and proportionate cost of your substantial happiness?

RECIPROCAL QUESTION B

Would I accept your comparable unhappiness if that were the unavoidable and proportionate cost of preventing my anger? Acceptance of the first sacrifice and rejection of the second indicates that the other person's happiness has greater reflectively endorsed priority. Acceptance of the second and rejection of the first indicates that avoidance of anger has greater priority. If the answer changes as the intensity or duration changes, the correct result is a threshold, not an absolute ranking. The word "legitimate" matters. Another person's happiness does not

automatically override rights, safety, fidelity, or duties to third parties. The test must therefore examine the actual commitments in conflict rather than treating happiness as an unrestricted good.

THE RECIPROCAL SACRIFICE TEST

- 4. Where the test functions better than observation alone
 - ⌘ Before action: It can evaluate unprecedented or future conflicts for which no decisive behavior yet exists.
 - ⌘ Where real experimentation would be unethical: It can examine dangerous, irreversible, or harmful choices without deliberately creating the harm.
 - ⌘ Where behavior is ambiguous: It can separate the commitments contained within a chosen package and ask which one the agent meant to protect.
 - ⌘ Where choice is constrained: It can distinguish endorsement from action produced by coercion, restricted options, fear, or necessity.
 - ⌘ Where impulse and identity diverge: It can distinguish what motivationally wins from what the agent reflectively authorizes to govern.
 - ⌘ Where thresholds matter: It can deliberately vary severity, probability, duration, and reversibility rather than waiting for many naturally occurring cases.
 - ⌘ Where moral interpretation matters: It can distinguish identical outward actions performed from loyalty, duty, fear, guilt, or another hidden commitment.

LIMIT

The test is vulnerable to dishonesty, self-deception, idealized self-description, and failure to imagine a real cost. It therefore becomes substantially stronger when later compared with actual conduct.

- 5. Revealed preference and Paul A. Samuelson

Paul A. Samuelson introduced the approach that became revealed preference theory in his 1938 paper "A Note on the Pure Theory of Consumer's Behaviour." His original language was "selected over." He developed the approach under the name "revealed preference" in 1948. Later economists, including Hendrik Houthakker and Hal Varian, extended its formal and empirical development. In its core form, revealed preference infers an ordering from observed choice: when an agent selects one available option while another was also feasible, the selection supplies evidence that the chosen option was revealed preferred in that choice situation. Formal revealed-preference

theory is a family of economic approaches rather than one single universal psychological test.

SAMUELSONIAN INSIGHT

What an agent actually selects among genuinely available alternatives provides evidence of the priority that was behaviorally effective in that choice. What observed choice does especially well

- ⌘ It confronts the agent with actual cost rather than cheap or idealized speech.
- ⌘ Repeated comparable choices can reveal a stable behavioral pattern.
- ⌘ It can expose self-misunderstanding when sincere predictions repeatedly conflict with conduct.
- ⌘ It supports empirical tests of consistency when feasible alternatives and constraints are known.

THE RECIPROCAL SACRIFICE TEST

What it cannot establish by itself

- ⌘ Why the option was chosen when several motives or commitments support the same action.
- ⌘ Whether the action expressed endorsement,

compulsion, habit, fear, coercion, or limited opportunity. * Whether the agent's choice promoted the agent's welfare. * Whether the behavior was morally justified. * What the agent would endorse before a choice that has never occurred. The combined method therefore uses the core revealed-preference insight as behavioral evidence without assuming that behavior alone transparently discloses the agent's deepest value, welfare, or moral judgment. Relation to earlier sacrifice-based reasoning

The general thought that sacrifice can indicate value ranking has prior philosophical use. Keith R. Peterson, for example, describes a "sacrifice test" in *A World Not Made for Us* (2020): what a person gives up to realize a higher value indicates a rank ordering. The distinct contribution proposed here is the controlled reciprocal reversal of the sacrifice, threshold mapping, framing checks, behavioral verification, and separation of endorsement from justification.

- 6. The Reciprocal Priority Verification Protocol

The strongest use of the Reciprocal Sacrifice Test is not to replace behavioral evidence but to combine controlled reflection with behavior and moral evaluation. The resulting method is the Reciprocal Priority Verification Protocol.

STAGE I - PROFESSED COMMITMENTS

Record what the agent claims to prioritize before interpreting behavior. Define A and B precisely and identify any relevant third commitments.

STAGE II - RECIPROCAL CLARIFICATION

Administer the Reciprocal Sacrifice Test, reverse the comparison, neutralize framing, and locate the threshold. Output: reflectively endorsed priority.

STAGE III - SAMUELSONIAN BEHAVIORAL VERIFICATION

Examine naturally occurring choices. Determine what alternatives were genuinely feasible, what the agent believed, what costs were expected, and whether the pattern is repeated. Output: behaviorally effective priority.

STAGE IV - DISCREPANCY ANALYSIS

Compare endorsement with conduct. Investigate coercion, weakness of will, self-deception, habit, addiction, incomplete information, changing values, unequal stakes, and hidden commitments.

THE RECIPROCAL SACRIFICE TEST

STAGE V - NORMATIVE ADJUDICATION

Use independent ethical reasoning to determine what ought to govern, considering rights, duties, justice, harm, consent, consequences, character, and third-party interests. Output: morally justified priority. The four-level result

Level

Question answered Professed priority What the agent says they value. Reflectively endorsed priority What survives reciprocal sacrifice under controlled reflection. Behaviorally effective priority What actually governs relevant conduct. Morally justified priority What ought to govern after independent normative analysis. Interpreting convergence and conflict * When all four

levels support A, the case displays unusually strong coherence and evidential convergence.

⌘ When endorsed priority favors A but effective priority favors B, the result may indicate weakness of will, coercion, hidden motivation, or idealized self-description. ⌘ When effective priority favors A but justified priority favors B, the protocol identifies a moral

failure without confusing that failure with factual uncertainty about what governed. ⌘ When professed and endorsed priorities differ, the agent may have discovered a deeper priority, revised a public claim, or exposed insincerity.

- 7. Why the combination is stronger than either method alone

The two methods are complementary because they examine different objects. The Reciprocal Sacrifice Test investigates controlled reflective endorsement. Revealed preference investigates choice patterns among feasible alternatives. Neither alone settles moral justification. Their combination prevents a common philosophical mistake: treating what a person says, what a person does, and what a person ought to do as though they were the same fact.

Path

Best feature Main limitation

Score

Direct ranking or introspection Fast statement of priority Weak control of framing, trade-offs, and self- deception 6.2 Revealed preference alone Strong evidence of effective choice Choice may underdetermine motive, endorsement, or justification 8.7 Reflective equilibrium Broad normative coherence among judgments and principles Not designed specifically to diagnose a two- commitment endorsement-behavior gap 8.3

THE RECIPROCAL SACRIFICE TEST

Path

Best feature Main limitation

Score

Multicriteria decision analysis Powerful for many criteria and quantified trade-offs Can overformalize morally qualitative commitments and does not itself separate endorsement from conduct 8.4 Reciprocal Sacrifice Test alone Controlled, prospective, two- direction priority clarification Hypothetical response may fail under real cost 8.9

Reciprocal Priority Verification

Protocol

Integrates profession, reciprocal endorsement, actual conduct, discrepancy diagnosis, and moral judgment More demanding; empirical superiority still requires testing 9.6 Scores are reasoned comparative judgments for one narrow task: diagnosing the full structure of a conflict between two commitments. They are not empirical benchmark results or rankings of each method's total importance in philosophy.

- 8. Philosophical uses

⌘ Moral disagreement: Identify whether disputants disagree about facts, about which commitments are in conflict, or about their priority threshold. ⌘ Thought experiments: Require reciprocal presentation of sacrifice and expose one-sided framing. ⌘ Akrasia and self-deception: Separate endorsed commitment from the motivation that defeats it. ⌘ Institutional ethics:

Compare an institution's stated values with the values preserved when compliance becomes costly. * Relationships and mediation: Clarify whether concern, autonomy, comfort, loyalty, safety, or control actually governs a conflict. * Public policy: State explicitly what liberty, equality, safety, welfare, or transparency a policy is

willing to sacrifice. * AI value evaluation: Compare a system's stated principles, controlled conflict responses, and behavior across repeated cases.

- 9. Limitations and safeguards

* The test does not prove moral correctness. Priority and justification remain separate.

* Perfectly matched sacrifices may be impossible; uncertainty should be reported rather than concealed. * A two-commitment framing may omit a third commitment that actually governs the case. * Hypothetical answers may be dishonest, self-deceived, emotionally unstable, or imaginatively weak. * Observed behavior may reflect constraint, coercion, ignorance, impairment, or lack of alternatives rather than endorsement. * Thresholds may change over time as experience, identity, and circumstances change. * Claims of superiority should remain limited to the specific diagnostic task and should be tested

empirically.

- 10. Conclusion

The Reciprocal Sacrifice Test makes a particular kind of hidden structure visible. It does not merely ask which option a person prefers. It asks which commitment the person reflectively

THE RECIPROCAL SACRIFICE TEST

accepts losing so that the other may govern, then reverses the sacrifice under controlled conditions. This yields a clearer account of endorsed priority and of the threshold at which that priority changes. Samuelson's revealed-preference insight adds a different and indispensable form of evidence: what the agent actually selected among genuinely available alternatives. When the reflective and behavioral results converge, confidence increases. When they diverge, the divergence becomes the object of inquiry rather than an inconvenience to be ignored. The Reciprocal Priority Verification Protocol unites these methods with discrepancy analysis and independent moral adjudication. Its central advantage is therefore completeness across four levels: what is professed, what is reflectively endorsed, what is behaviorally effective, and what is morally justified. For the narrow task of diagnosing two conflicting commitments across reflection, action, and ethics, no examined alternative combines all of these functions in one procedure.

DEFENSIBLE STATUS

The protocol is a strongest-in-class conceptual candidate for its exact task. It is more comprehensive than any single compared pathway, but its predictive reliability and empirical superiority should be tested rather than assumed.

References

Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112 (4): 659-688. Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1): 5-20. Hands, D. Wade. 2014. "Paul Samuelson and Revealed Preference Theory." *History of Political Economy* 46 (1): 85- 116. Hansson, Sven Ove, and Till Grune-Yanoff. "Preferences." *Stanford Encyclopedia of Philosophy*. Peterson, Keith R. 2020. *A World Not Made for Us: Topics in Critical Environmental Philosophy*. SUNY Press. Samuelson, Paul A. 1938. "A Note on the Pure

Theory of Consumer's Behaviour." *Economica* 5 (17): 61-71. Samuelson, Paul A. 1948.
"Consumption Theory in Terms of Revealed Preference." *Economica* 15 (60): 243-253.
Sen, Amartya. 1973. "Behaviour and the Concept of Preference." *Economica* 40 (159): 241-259.
Varian, Hal R. 2006. "Revealed Preference." In *Samuelsonian Economics and the Twenty-First Century*.

FINAL COMPARATIVE RATING

9.6 / 10 Strongest integrated conceptual pathway presently identified for diagnosing a two-commitment conflict across reflection, behavior, discrepancy, and moral justification.
Qualification: conceptually strongest-in-class, not yet empirically proven universally superior.