

The Temporal Welfare Trade-off Framework: A Decision-Support Tool for Animal Behavior Modification and Related Welfare-Relevant Procedures

Roberto Barata

Human-Animal Science, Denmark

[Doi:10.19044/esj.2026.v22n27p1](https://doi.org/10.19044/esj.2026.v22n27p1)

Submitted: 19 June 2026

Accepted: 16 September 2026

Published: 30 September 2026

Copyright 2026 Author(s)

Under Creative Commons CC-BY 4.0

OPEN ACCESS

Cite As:

Barata, R. (2026). *The Temporal Welfare Trade-off Framework: A Decision-Support Tool for Animal Behavior Modification and Related Welfare-Relevant Procedures*. European Scientific Journal, ESJ, 22 (27), 1. <https://doi.org/10.19044/esj.2026.v22n27p1>

Abstract

Applied ethology and animal training lack a shared analytical procedure for evaluating procedures that impose short-term welfare costs to achieve longer-term welfare or survival outcomes. Existing instruments address parts of the problem, leaving intensity, duration, probability, and temporal discounting uncombined. The Temporal Welfare Trade-off Framework proposed here draws on cumulative welfare assessment in ethology, delay discounting in comparative psychology, and harm-benefit analysis in research ethics. A candidate procedure is expressed as a benefit-cost ratio: the hyperbolically discounted present value of an expected long-term outcome over a cost function combining intensity, duration, frequency, and reversibility. Its equations are a provisional scoring convention adopted for auditability. Since the inputs are ordinal and the combination rules are modeling choices, the output carries no interval or ratio-scale meaning. Three categorical thresholds bound the arithmetic: an intensity ceiling, a reversibility floor, and an alternative-exhaustion requirement, while an epistemic flag records where evidence is too thin for a confident verdict in either direction. The framework evaluates a procedure's welfare structure, setting its methodological label aside. Three worked cases, drawn from conservation translocation, working-dog management and companion-animal husbandry, run the framework on illustrative values throughout. Reliability testing,

parameter calibration, and break-even ratio calibration remain outstanding, so the framework currently ranks candidate procedures against one another.

Keywords: Animal welfare, harm-benefit analysis, temporal discounting, Five Domains Model, applied ethics, welfare audit, behavior modification

1. Introduction

The ethical evaluation of procedures that impose short-term welfare costs to achieve longer-term welfare or survival outcomes is a recurring problem in applied ethology, conservation behavior, and animal training. The framework developed here is offered as a theoretical proposal and a decision-support tool, advanced for further testing, and it awaits empirical validation, inter-rater reliability testing, and practical calibration before any use as a settled standard.

The framework proposed here is method-neutral. The same analytical procedure applies to on-field training, environmental modifications, pharmacological interventions, and surgical procedures. Each of these may impose short-term welfare costs in pursuit of a longer-term outcome, and each requires the same evidentiary scrutiny. A framework that singled out any one methodological category for stricter or more lenient evaluation would mistake the methodological label for the welfare-relevant variables: intensity, duration, frequency, reversibility, and the probability and magnitude of the expected outcome. The unit of analysis is therefore a specified protocol delivered to a specified individual in a specified context, and never a method class. Intensity is a property of delivery, so a class of methods has no intensity score and cannot enter the model. Claims that a class of tools is acceptable, or unacceptable, fall outside what this framework can evaluate, and a practitioner who holds either view can apply the framework within it and see where the welfare-relevant variables and the commitment diverge.

Welfare science has produced substantial conceptual and methodological progress relevant to this question. The Five Domains Model offers a structured account of the physical and affective components of welfare, with explicit attention to compromise and to enhancement (Mellor et al., 2020). The cumulative welfare assessment tradition, represented by the Animal Welfare Assessment Grid and the Extended Welfare Assessment Grid, supplies tools for tracking welfare states across lifetime events (Honess and Wolfensohn, 2010; Wolfensohn et al., 2015). The Welfare Footprint Framework quantifies pain and pleasure as functions of intensity and duration, providing welfare science with a common currency for comparison across systems (Schuck-Paim and Alonso, 2025). The life-worth-living threshold, introduced by the Farm Animal Welfare Council and critically examined by Yeates, provides a normative baseline below which no individual should be

allowed to fall (Farm Animal Welfare Council, 2009; Mellor, 2016; Yeates, 2011, 2012). The harm-benefit analysis required in research contexts offers a structured procedure for weighing costs against outcomes, although critics have highlighted its subjectivity and vulnerability to instrumental abuse (Brønstad et al., 2016; Grimm et al., 2019; Laber et al., 2016).

A parallel literature in comparative psychology and behavioral economics has mapped in extensive detail how individuals across many species weigh outcomes distributed over time. Hyperbolic discounting supplies a mathematical structure for time-sensitive valuation, fitted to choice data in pigeons, rats, and humans and traced to Chung and Herrnstein's (1967) finding that pigeons match relative responding to the reciprocal of reinforcement delay (Mazur, 1987; Vanderveldt et al., 2016). Preference reversal as a delayed outcome approaches makes temporal position a decision variable in its own right (Ainslie, 1992; Laibson, 1997). Ten rats choosing between a smaller unshocked alternative and a larger one carrying a delayed shock discounted the aversive consequence hyperbolically (Rodríguez et al., 2018). These findings await systematic incorporation into welfare-assessment frameworks, despite their direct bearing on the trade-off problem.

Welfare scientists have developed tools for describing cumulative welfare, applied ethicists have developed structured procedures for harm-benefit judgment, and comparative psychologists have described the temporal structure of valuation in sentient individuals. No published framework combines these three traditions into a single structure capable of addressing the practical question that practitioners, ethics committees, and conservation managers face, which is whether a proposed procedure, applied with specified intensity and duration, can be defended on the basis of the probability-weighted, time-discounted welfare or survival outcome it is expected to secure.

The framework proposed here is defended on four grounds. First, every variable in the model has an established measurement tradition in welfare or behavioral science to build on, which supplies anchored descriptions for scoring, although the validity of those components does not transfer to the composite, and the combined model remains unvalidated. Second, the discounting function has empirical support across taxa for its functional form, whatever uncertainty attaches to its parameters. Third, the framework is bounded by three categorical thresholds that prevent the utilitarian logic of the core equation from justifying severe or irreversible harm. Fourth, the framework supports an epistemic flag that identifies procedures whose evidence base is too thin for the model to issue a confident verdict, separating evidentiary inadequacy from substantive disagreement about welfare costs and outcomes.

The framework is an analytical tool that makes welfare trade-offs reviewable and harder to argue from intuition alone. The phrase decision-support tool, in the title and throughout, denotes an aid to structured comparison between the options a practitioner has. Whether any procedure, taken on its own, is ethically acceptable lies outside what the framework can establish and awaits the calibration described in Section 8. Its numerical output formalizes the assessor's reasoning. The practitioner retains judgment and responsibility for the decision.

2. Theoretical Foundations

2.1 Cumulative Welfare

Welfare science has moved from point-in-time assessments to cumulative measures that track welfare over an individual's lifetime. The view rests on the study of animal emotion, where short-term affective states accumulate into longer-term mood (Mendl et al., 2010), on the treatment of welfare as a measurable state of the individual in relation to its environment (Broom, 1991), and on the conception that ties welfare science to the ethical concerns motivating it (Fraser et al., 1997). Affective states integrate across time in ways single-snapshot measures cannot capture, and a procedure's ethical weight depends on how its effects accumulate, sensitize, habituate, or propagate beyond the immediate context (Bateson and Poirier, 2019; Browning, 2022). Which accumulations count as welfare costs depends in part on the conception of welfare in use, and the field sustains several at once (Weary and Robbins, 2019).

The Animal Welfare Assessment Grid developed by Wolfensohn and colleagues captures four parameters, namely physical, psychological, environmental, and procedural, and produces a Cumulative Welfare Assessment Score plotted across an individual's lifetime to reveal the welfare impact of discrete events (Wolfensohn et al., 2015). The Welfare Footprint Framework quantifies cumulative pain and pleasure as functions of intensity and duration using the Pain-Track and Pleasure-Track notation systems (Schuck-Paim and Alonso, 2025). Recent work on positive welfare assessment in dairy cattle has argued that a protocol should distinguish indicators by the temporal scale of the affective state being measured, with short-term emotion, medium-term mood, and longer-term cumulative affect each requiring different indicator sets, a temporal decomposition that corresponds to the structure proposed below (Keeling et al., 2021).

These cumulative frameworks share three commitments that the present model inherits. Welfare is a property that accrues across time, so a single-moment assessment captures only a fragment of what matters morally. Both negative and positive affective states contribute to the cumulative balance, so a procedure imposing short-term negative affect must be evaluated

against the affective profile it is expected to produce later. Intensity and duration are the proximate measurable variables, with frequency, reversibility, and severity as modifiers determining how episodes combine into a cumulative burden.

The limit of these frameworks for present purposes concerns their handling of future outcomes, since cumulative welfare scores are applied retrospectively or as running measures while the problem a practitioner faces is prospective. The present model draws its variables from that tradition while leaving aside its integration method, computing a within-episode cost and a discounted outcome in place of a welfare trajectory. Defending whether a future procedure can be justified by a future welfare or survival outcome requires an additional analytical tool: discounting future outcomes.

2.2 Temporal Discounting

A substantial literature in comparative psychology, beginning with Chung and Herrnstein's (1967) pigeon experiments on delay of reinforcement and extended by Ainslie, Mazur and others, has established that pigeons, rats and humans weigh outcomes according to a hyperbolic discount function (Ainslie, 1992; Mazur, 1987; Vanderveldt et al., 2016). The form used here, $V = A / (1 + kD)$, where V is the present subjective value, A is the outcome magnitude, k is a species-specific and individual-specific discount parameter, and D is the delay, comes from Mazur (1987) and fits behavioral data better than the exponential form of classical economic theory. Comparative reviews prefer its hyperboloid generalization, which adds an exponent to the denominator (Vanderveldt et al., 2016).

Negative outcomes are discounted too, broadly along the same hyperbolic curve (Rodríguez et al., 2018). Nor is discounting confined to the future: human participants discount past outcomes as well (Yi et al., 2006), so the function tracks temporal distance instead of futurity. Whether the sign of an outcome modulates the rate, and whether any such modulation generalizes across species, is unsettled. The framework treats that as an open parameter, while taking from the well-documented species-specific estimates the possibility of calibrating the function from existing data.

Range and construct both bound what the framework can take from that literature, and both are consequential enough to belong here instead of among the limitations. Delay-discounting parameters are estimated in choice procedures operating over seconds to minutes in non-human animals, while the horizons over which welfare outcomes accrue in the applications considered here run to hundreds or thousands of days. Applying a function fitted over the former to the latter is extrapolation beyond the fitted range, and in the hyperbolic family the implied instantaneous rate falls as delay grows, so the extrapolation is not neutral with respect to the conclusion it supports.

Delay discounting also measures revealed preference at a moment of choice, and a choice rate is not the welfare subsequently experienced, so using the first to weight the second imports an animal's impulsivity into an ethical calculus. The framework treats the two as distinct, uses the discount to represent the individual's temporal valuation, leaves the moral weight of distant welfare with the assessor, and allows setting the rate to zero when an assessor judges that a decision should give full, undiscounted weight to the individual's future. Any application should report the discounted and undiscounted comparison together so a reader can see how much the discount changes the result. Where the two diverge enough to change an ordering, the assessor must justify relying on the discounted figure, since an individual with a steeper discount function would otherwise have its future counted for less on the strength of its own impulsivity, which no one would defend as a reason.

The ethical implications of hyperbolic discounting for welfare science are less frequently discussed. A framework weighing short-term costs against longer-term outcomes must reflect the structure of temporal valuation, or it will misvalue deferred outcomes, while a discount that is too aggressive inflates present cost and makes welfare-justified procedures impossible to defend. Both errors are possible, and neither can be excluded at present, which is why Section 3.3 requires the rate, its source and a sensitivity analysis to be reported, and why the rates used here are provisional illustrative inputs awaiting species-specific measurement.

2.3 Harm-Benefit Analysis

Directive 2010/63/EU requires harm-benefit analysis as part of the ethical evaluation of animal research protocols (European Parliament and Council of the European Union, 2010), and comparable structures operate under the United Kingdom's Animals (Scientific Procedures) Act and the United States Animal Welfare Act through Institutional Animal Care and Use Committees. The AALAS-FELASA Working Group produced influential guidance on structuring harm-benefit analysis in practice, and a parallel literature has examined how to incorporate reproducibility, experimental validity, and the 3Rs into harm-benefit reasoning (Brønstad et al., 2016; Laber et al., 2016; Würbel, 2017).

Critics of harm-benefit analysis, most pointedly Grimm et al. (2019), have argued that the procedure is too subjective to produce consistent judgments, too vulnerable to optimistic estimates of benefit, and too easily deployed as ethical cover for procedures that would not survive rigorous evaluation of either the harm or the outcome in isolation. The critique deserves serious attention, because any framework following the harm-benefit logic inherits its vulnerabilities. The response embedded here is that the utilitarian logic should be bounded by categorical thresholds for the most consequential

parameters, and that an epistemic flag should separate procedures with adequate evidentiary support from those the evidence cannot settle. These features address the most common abuse of harm-benefit analysis: justifying severe harm through inflated estimates of future outcomes under conditions of evidential thinness.

2.4 Training Ethics Frameworks and the Empirical Literature

The Least Intrusive, Minimally Aversive (LIMA) principle, introduced by Lindsay (2005) and later adopted and adapted by professional dog-training associations and extended to other companion and working species, requires the use of the least intrusive, minimally aversive technique reasonably expected to succeed, combined with a competence criterion specifying that only adequately trained practitioners should make the assessment. The associated Humane Hierarchy of Behavior Change, developed by Friedman (2009), orders interventions from wellness and antecedent arrangement through positive reinforcement, differential reinforcement, extinction, and negative reinforcement, and finally positive punishment as a last resort. Both are professional statements, cited here for the normative commitments they express. Neither supplies efficacy evidence, and the framework maintains that distinction wherever sources of this kind appear.

The more recent LIFE model proposed by Fernandez (2024) critiques LIMA for its ambiguous terminology, historical misuse to justify aversive methods, and insufficient attention to the learner's choice and functional assessment. LIFE specifies three commitments: increasing choices by inhibiting less, emphasizing functional assessment over symptom suppression, and defining training success beyond behavioral effectiveness alone.

These frameworks supply much of the qualitative scaffolding the present model uses as its analytical baseline. The LIMA requirement that less aversive alternatives be exhausted corresponds to the Alternative Exhaustion Threshold, the Humane Hierarchy's ordering of interventions matches the intensity grading proposed for the cost function, and the LIFE model's commitment to functional assessment and learner choice aligns with the reversibility and probability parameters.

Several frequently cited primary studies have design features that constrain what can be inferred, and comparable features appear on both sides of the debate. Vieira de Castro et al. (2020) recruited dogs through the training schools their owners had already chosen, an observational structure in which training method and owner characteristics covary. Samples are generally small, and Johnson and Wynne (2024) allocated 19 of the 30 dogs recruited across three conditions. Cooper et al. (2014) and Vieira de Castro et al. (2020) paired salivary cortisol with behavioral coding, an assay that Cobb et al. (2016) show, through systematic review and meta-analysis, is an imprecise

proxy in dogs. China et al. (2020) worked from behavioral outcome measures alone, and Sargisson and McLean (2021) discuss in commentary the inferences those measures support.

These observations bear on the quality of the evidence alone, and they leave intact the substantive welfare concerns the studies raise. Practitioner beliefs influence how a given behavior is read, since training-style commitments can predict whether a problem is treated as a clinical disorder or a manageable training matter (Hunter et al., 2020). A framework that asks practitioners to record their reasoning makes those commitments visible, where they can be examined openly.

What these frameworks and the literature behind them leave open, and what the present model supplies, is a method-neutral quantitative temporal structure that allows comparison of any specific proposed procedure against any specific expected outcome under explicit epistemic conditions, with the same evidentiary standards applied regardless of the methodological category from which the procedure is drawn.

3. The Framework

The equations below are a scoring convention, adopted because they can be audited. Intensity, magnitude and reversibility are judgments made against anchored descriptions taken from the Five Domains Model and from Annex VIII of Directive 2010/63/EU, and those anchors are what the scales rest on. They are not bare ordinals, since a bare ordinal admits any order-preserving relabelling, and a relabelling that asserted the step from four to five to be twenty times the step from two to three would contradict the descriptions. Neither are they interval scales, since nothing establishes equal steps. Treating them as anchored and approximately spaced is the standing practice in severity classification, and it is what species-specific rubric calibration would test.

A restriction on what the arithmetic asserts follows from that. Multiplying an intensity of 4 in place of 2 records a judgment that one procedure is more intense, and asserts nothing about how much. The framework therefore claims dominance relations, an ordering that survives every plausible rescaling of the inputs, and claims nothing where plausible ranges overlap. Reporting no ordering is a correct output and will be the common one. Krueger et al. (2023) had 97 raters classify photographed mice by severity, 84 of them entering the final analysis in three groups by experience, and found severity accounting for 91.6 percent of variance at the extremes while contributing nothing at mild-to-moderate levels, where rater interactions dominated, and experience accounted for none of it. Distinctions between adjacent middle grades are accordingly treated as unresolved unless a case shows otherwise.

Empirical propositions, validated relationships, modeling conventions, normative assumptions, and illustrative values are combined in what follows, and they should be kept apart. An empirical proposition is a claim about the world supported by cited evidence, while a validated relationship is a functional form established by data. The framework contains one of the latter, the hyperbolic shape of delay discounting. Modeling conventions are structural choices made for tractability and transparency, such as the multiplicative combination of the cost variables, and normative assumptions are ethical commitments, such as the position of the intensity ceiling. Illustrative values demonstrate the machinery. None is an empirical estimate. Table 1 classifies every parameter and functional form into those categories, and that classification governs how much weight each component can bear.

Table 1: Epistemic status of each component of the framework.

Component	Class	Basis	What calibration would settle
Intensity, I, 1 to 5	Empirical proposition, scored against anchors	Anchored severity descriptions, Five Domains and Annex VIII	Species-specific rubrics and inter-rater reliability
Duration, D	Empirical proposition	Observed exposure per episode	Which portion of an episode is welfare-relevant
Frequency, F	Empirical proposition	Counted episodes	Whether sensitization or habituation predominates
Reversibility, R, 0 to 1	Empirical proposition, scored against anchors	Conditioned fear, learned helplessness and habituation timelines	Species-specific recovery data
Amplifier, 2 - R	Modeling convention, normatively motivated	Linear penalty expressing the weight placed on lasting harm	Whether the penalty is linear and doubles at the limit
Cost form, $I \times D \times F \times (2 - R)$	Modeling convention	Each term scales the burden and none is dispensable	Whether another monotonic combination fits judgment better
Magnitude, M, 1 to 5	Empirical proposition, scored against anchors	Same anchored scale as intensity	Species-specific rubrics
Probability, P, 0 to 1	Empirical where efficacy data exist, otherwise illustrative	Published efficacy for the technique, species and context	Efficacy trials in the relevant domain
Expected duration, T	Empirical proposition	Actuarial expectation for the post-intervention context	Species and context-specific survival data
Outcome form, $M \times P \times T$	Modeling convention	Risk-neutral expected value, dimensions treated as independent	Whether low-probability severe outcomes need a risk adjustment
Hyperbolic discount form	Validated relationship	Delay-discounting studies across taxa	Nothing on form. Range and construct remain open
Discount rate	Illustrative value	Stipulated and small, with no species estimate for dogs	Delay-discounting parameters at the horizons used here

Component	Class	Basis	What calibration would settle
Intensity ceiling at I = 5	Normative assumption	Severe category of Annex VIII	Nothing empirical. A normative choice open to revision
Reversibility floor at R = 0.4	Normative assumption	Point at which lasting cross-domain compromise becomes probable	Evidence on where that point falls
Alternative exhaustion	Normative assumption	LIMA, Humane Hierarchy and ALARA ordering	Nothing empirical. An evidentiary requirement
Break-even benefit-cost ratio	Not yet specified	No value is fixed by the arithmetic	Expert-consensus calibration. Until then, no isolated verdict

Note: Illustrative values demonstrate the framework and are not empirical estimates. Modeling conventions and normative assumptions are choices, open to substitution by an assessor who states the substitute and applies it consistently.

3.1 The Central Equation

The framework scores a candidate procedure by computing a Net Welfare Value, the hyperbolically discounted expected welfare outcome minus the structured short-term welfare cost. A procedure satisfies the framework's analytical conditions when its benefit-cost ratio is favorable against the alternatives considered, the three categorical thresholds are satisfied, and the alternative-exhaustion documentation withstands independent review. The scoring convention takes the following form.

$$NWV = PV(B) - C$$

PV(B) is the present value of the expected long-term welfare outcome, calculated using a hyperbolic discount function, calibrated to the species and context where data allow and otherwise stipulated and reported as such. C is the structured short-term welfare cost, calculated from intensity, duration, frequency, and reversibility parameters. Treating these parameters as indices of a real burden on the individual follows the established account of the stress response as a measurable physiological and behavioral cost (Moberg, 2000). The three categorical thresholds, specified in Section 3.4, function as gates that the arithmetic cannot override.

The equation is kept simple by design, because the epistemic reality of welfare assessment supports only limited precision, well short of what a more elaborate formalism would imply. Beausoleil and Mellor (2015) emphasize that grades of this kind are qualitative and support only relative rankings on a common scale, and that caution applies to the present model. Where the assessor's confidence in any component falls below the level specified in Section 3.5, the framework records an epistemic flag and reports the result as

unresolved. The arithmetic's present role should also be stated plainly. Until the break-even ratio is calibrated, the ratio orders the alternatives an assessor must consider and identifies which one the alternative-exhaustion threshold points to, and the categorical gates determine the outcome. Section 6.2 shows that sequence at work, with the case settled at Threshold 3 before consulting the arithmetic.

The cost term and the outcome term rely on duration measures that need not share the same physical unit, which affects how to interpret the output. The cost is computed from a within-episode duration, typically seconds or minutes, while the outcome is computed from a duration measured across the individual's remaining life, typically days or years, so changing the unit for either term rescales one side of the subtraction and leaves the other untouched. The sign of the Net Welfare Value is therefore an artifact of the convention in use and cannot serve as a decision criterion on its own.

The framework accordingly takes the benefit-cost ratio, the discounted outcome divided by the cost, as its decision-relevant statistic. A change of unit rescales that ratio by the same constant for every option under comparison and leaves their ordering intact, provided the discount rate and the delay are expressed in matched units, since options with different delays are otherwise rescaled by different amounts. The Net Welfare Value is retained as the audit record, because it exposes the two quantities separately and shows a later reader where two assessors part company.

Where the break-even value of the ratio falls is not fixed by the arithmetic and depends on the same unit convention, so the empirical calibration described in Section 8 must establish it before the framework can classify a procedure considered in isolation. Until that calibration exists, the framework ranks the alternatives available in a given case, which is the comparative question a practitioner actually faces, and any application should report the duration convention it adopted so that another assessor can reproduce the ranking and challenge it on its own terms.

3.2 The Short-Term Cost

The cost function decomposes the procedure's welfare impact into four scored variables, with reversibility as a multiplicative amplifier.

$$C = I \times D \times F \times (2 - R)$$

Intensity is scored on a one-to-five scale using observable behavioral and physiological indicators drawn from the Five Domains Model's mental-state domain, calibrated to species. One corresponds to mild, transient discomfort, five to severe distress, consistent with the severe category of Annex VIII of Directive 2010/63/EU and used in FELASA guidance

(European Parliament and Council of the European Union, 2010). Intermediate values reflect graded behavioral indicators validated in species-specific welfare literature, including conflict behaviors, displacement activities, tonic immobility, and stress-related physiological markers (Mellor et al., 2020). Inference of an affective state from any such indicator remains contested (Paul et al., 2020). For dogs, grading should use cognitive-bias methodologies (Mendl et al., 2009), and physiological indicators should be evaluated critically (Cobb et al., 2016). Species-specific rubrics should be compiled as an annex to any operational deployment.

Duration is expressed in standardized time units, typically minutes or seconds of welfare-relevant exposure per episode. Consistency across comparisons matters more than the unit chosen. The variable captures the within-episode extent of the welfare-relevant experience, while frequency captures cumulative exposure across the intervention period.

Frequency is the number of welfare-relevant exposures across the intervention period. Recurrence increases cumulative cost through summation and, where the relevant mechanisms are active, through sensitization. Where habituation predominates, the multiplier can be attenuated, with the assessor justifying the choice from species-specific evidence.

Reversibility is scored on a zero-to-one scale, 1 representing fully reversible affective and behavioral impact and 0 representing permanent or deeply conditioned effects. The expression $(2 - R)$ amplifies the cost, leaving a fully reversible procedure unchanged and doubling an irreversible one, which reflects the ethical weight placed on irreversible harm in welfare legislation, bioethics, and harm-benefit analysis (Brønstad et al., 2016; European Parliament and Council of the European Union, 2010; Laber et al., 2016). Scoring should use the literature on conditioned fear, learned helplessness, single-event learning, and habituation timelines, with the empirical basis specified.

3.3 The Long-Term Outcome and its Present Value

The outcome calculation parallels the cost calculation in structure, with magnitude, probability, and expected duration of outcome scored as the three primary variables.

$$B = M \times P \times T$$

Magnitude is scored on the same one-to-five scale as intensity, applied to the expected welfare gain or to the reduction in expected welfare loss. A welfare outcome may combine preventing future negative affect, creating future positive affect, and reducing the probability of severe adverse outcomes such as mortality. For anti-predator training of a translocation candidate,

magnitude combines the reduction in predation probability with the welfare implications of continued survival and species-typical behavioral expression.

Probability is expressed as a value between 0 and 1, representing the probability that the intervention produces the intended outcome. The variable carries substantial weight, as it brings operator competence, species-typical learning parameters, and the evidence base for the specific technique into the model. For interventions with published efficacy data of adequate methodological quality, take probability from the relevant studies, with attention to the cautions in Section 2.4. For motivational domains where conventional procedures are mechanistically mismatched, as addressed in Section 4, set probability low on the reading developed there, whatever the category, and record an epistemic flag because that reading is a mechanistic prediction awaiting testing. Absent efficacy data, the assessor specifies an estimate from available indirect evidence and documents the inferential basis, flagging the procedure under Section 3.5.

Expected Duration of Outcome is specified in the same time units used for the delay. A welfare gain persisting for years carries more moral weight than one dissipating within weeks, even at comparable peak magnitude. T is constrained by the expected remaining lifespan under plausible post-intervention conditions, with an explicit upper bound that the assessor justifies from the species-specific actuarial literature.

The present value of the outcome is calculated using a hyperbolic discount function.

$$PV(B) = B / (1 + \delta \times t)$$

δ (delta) is a discount rate, drawn from the comparative psychology literature for the species in question where such data exist, and t is the temporal distance between the welfare-relevant intervention and the realization of the outcome. Leonardi et al. (2012) measured tolerance of delay in an exchange task, which is a different quantity from a discount parameter, so the value of δ used for dogs in this paper is a stipulated illustrative input with no species-specific estimate behind it, while Vanderveldt et al. (2016) supply the cross-species evidence for the form of the function in pigeons, rats, and humans instead of direct data on dogs. Comparable literature exists for horses, rats, pigeons, and several primate species, and reasonable interpolation is possible for closely related species where direct data are absent. The discount rate should be reported transparently in any application of the framework to enable sensitivity analysis. Where the species-specific discount rate is unavailable, the assessor must report the rate used, the source from which it was drawn, and the sensitivity of the verdict to alternative parameterizations.

The discount is applied to the delay between the procedure and the onset of the outcome, not across the period over which the outcome accrues,

so a benefit persisting for years is treated as a quantity arriving at one point. An assessor who prefers to discount the stream can integrate the same function across the accrual period, giving an effective duration of $(1/\delta) \ln[(1 + \delta(t + T)) / (1 + \delta t)]$ in place of T , a substitution that shortens the effective duration in each of the three worked cases and changes no ranking.

The choice of hyperbolic over exponential discounting follows the empirical finding that sentient individuals across species discount hyperbolically, and the form of the function itself. The hyperbolic decays as the reciprocal of delay while the exponential decays geometrically, so for two functions matched at a common delay the hyperbolic declines more steeply over short horizons and retains more weight over long ones. That combination produces the preference reversals the comparative literature reports.

3.4 Categorical Thresholds

The arithmetic of the Net Welfare Value equation, taken alone, would permit any cost to be justified by a sufficiently large or sufficiently probable outcome. That is the fundamental vulnerability of utilitarian welfare models, and it is why critics of harm-benefit analysis have warned against its unqualified deployment (Grimm et al., 2019). The framework addresses this vulnerability by embedding three categorical thresholds that serve as gates, such that failure at any threshold disqualifies the procedure regardless of the computed values. The evidence bearing on each threshold is recorded before any scoring, since an assessor cannot document the exhaustion of alternatives after committing to a procedure, and the three verdicts are entered after the arithmetic, since alternative exhaustion cannot be adjudicated until the alternatives have themselves been scored. The assessment sheet in Section 5.1 follows this order.

Threshold 1: Intensity Ceiling

No procedure scoring $I = 5$ passes this threshold, whatever its computed values. The ceiling is a normative commitment, adopted on the principle that the framework should not function as a permission system for severe suffering, regardless of the projected outcome or the intervention's methodological category. A procedure causing severe distress, consistent with the severe category of Annex VIII of Directive 2010/63/EU, requires a more demanding ethical procedure than this framework provides, and its evaluation belongs in a separate institutional process with explicit ethics committee oversight. A score of $I = 4$ can be evaluated through the framework but requires the most rigorous documentation of alternative exhaustion and the strongest evidentiary support for the projected outcome (Beauchamp and DeGrazia, 2020).

Threshold 2: Reversibility Floor

Any procedure scoring R below 0.4 is refused, with the verdict subject to second-assessor review in a formal institutional context. The floor is likewise a normative commitment, reflecting the documented risk that chronic welfare compromises propagate across life domains in ways that cannot be undone (Bateson, 2016; Bateson and Poirier, 2019). Surgical interventions such as devocalization, which permanently remove communicative behaviors while leaving the motivational substrate untouched, fail at this threshold for the same reasons that chronic pharmacological suppression of behavior would. A reversibility score below 0.4 indicates lasting effects whose ethical weight is too substantial for a simplified audit.

Threshold 3: Alternative Exhaustion

The framework refuses any procedure that has not cleared a documented exhaustion of less invasive alternatives, in the spirit of the LIMA and Humane Hierarchy ordering and analogous to the As Low As Reasonably Achievable principle in radiation medicine. Documentation must include the protocols attempted, the durations and reinforcement schedules used, the environmental and management interventions tried, the practitioner competence under which alternatives were trialed, and the measurable outcomes recorded. This documentary-evidence requirement is met by independently observable evidence, and any application that cannot produce it is refused here without further calculation (Fernandez, 2024; Friedman, 2009). Alternatives that meet the requirement have adequate documented efficacy, so an available but unevidenced option does not, by itself, disqualify a procedure.

The threshold operates symmetrically across methodological categories. A pharmacological intervention proposed for behavioral management requires documentation that environmental, behavioral, and antecedent-arrangement alternatives have been exhausted, and a surgical intervention requires documentation that pharmacological, behavioral, and environmental alternatives have been exhausted.

3.5 The Epistemic Flag

Beyond the three categorical thresholds, the framework supports an epistemic flag for procedures the evidence cannot settle either way. The mechanism operates as a flag because evidentiary thinness, by itself, does not constitute a substantive welfare argument against the procedure. The flag distinguishes three conditions.

The first condition concerns the probability of the outcome. Where probability is estimated without published efficacy data of adequate methodological quality for the specific technique, species, and context, the

procedure carries a flag indicating that the variable rests on indirect evidence or practitioner extrapolation. The flag is informational and leaves the procedure eligible, with the verdict contingent on probability assumptions a reviewer should examine.

The second condition concerns the discount rate. Where the species-specific discount rate is interpolated from a phylogenetically distant relative, with direct empirical data on the species unavailable, the procedure carries an epistemic flag indicating that δ rests on inferential calibration drawn from related species. The flag is informational, and the assessor should report a sensitivity analysis showing how the verdict varies under alternative δ values.

The third condition concerns the alternative-exhaustion documentation. Where the documentation rests on practitioner self-certification and independently observable evidence of attempted alternatives is unavailable, the procedure carries an epistemic flag indicating that the Threshold 3 verdict depends on the assessor's trust in the practitioner's account.

The cumulative effect of these flags is to make the framework's epistemic state explicit at the point of decision and visible in the numerical output. A procedure with a favorable ratio, all three categorical thresholds satisfied, and no epistemic flags raised represents the strongest case the framework can make. A procedure with a favorable ratio but multiple flags represents a case where the analytical structure has been satisfied while the evidentiary basis remains contestable, and the assessor's report should present the verdict as provisional.

The three categorical thresholds and the epistemic flag together address the strongest critique of utilitarian welfare models, asserting that certain constraints on permissible action stand outside any trade against projected outcome, and that the absence of evidence bears on what can be concluded.

3.6 Calibration and the Status of the Parameters

The intensity and reversibility scales are graded in discrete steps because welfare science already grades severity that way, with the Five Domains Model and Annex VIII of Directive 2010/63/EU supplying anchored descriptions at each level (European Parliament and Council of the European Union, 2010; Mellor et al., 2020). The reversibility term takes the form two minus R so that a fully reversible procedure leaves the cost unchanged while an effectively irreversible one doubles it, expressing in a single linear penalty the heavier ethical weight welfare legislation and bioethics place on lasting harm (Bateson, 2016).

The cost function is linear in intensity, which treats intensity and duration as compensatory within the range the framework evaluates. An

assessor who holds that intensity should carry more weight can substitute $C = I^\alpha \times D \times F \times (2 - R)$ and report α , with $\alpha = 1$ adopted here as the transparent default. Applied to the second worked case, raising α lowers each option's ratio in proportion to its intensity, so the high-intensity scenario falls fastest and drops below the milder aversive scenario at approximately $\alpha = 1.45$. Setting $\alpha = 1$ therefore gives the high-intensity option its most favorable reading, which makes it the parameterization least favorable to the conclusion this paper reaches.

The multiplicative structure of both terms is likewise a design choice. The cost variables combine multiplicatively because the structure expresses the judgment that intensity, duration and frequency each scale the burden, and that a procedure approaching zero on any one of them imposes little cost, which additive alternatives do not capture. The outcome variables combine multiplicatively because the product is the expected value of the outcome, which assumes risk neutrality and treats magnitude, probability and duration as independent. Neither form was derived from data, no welfare evidence establishes either as correct, and each can be replaced by another monotonic form provided the substitution is stated and applied to every option under comparison.

The categorical thresholds sit where the severity literature already draws its lines: the intensity ceiling at the level the Directive classifies as severe, and the reversibility floor where lasting cross-domain compromise becomes probable. The hyperbolic form of the discount follows the shape comparative delay-discounting studies report across species, and the size of δ is an illustrative starting value chosen to be small, a provisional input the assessor should select and disclose case by case while measurement remains outstanding. Assigning severity remains a human judgment in every welfare instrument in current use, from the Animal Welfare Assessment Grid to the Directive's own severity classification, because no validated cross-species metric of welfare intensity exists to replace it. The framework inherits that condition instead of creating it, and its response is to record where each judgment entered and on what basis, so that a reader can substitute their own and see the consequence. Calibration against expert consensus, inter-rater reliability testing, and species-specific rubrics remain necessary before the parameters can be treated as settled.

4. Predation as a Separate Domain

The framework requires a separate treatment of predation because predation occupies a motivationally distinct domain in which conventional therapeutic procedures based on counterconditioning and desensitization are, on the mechanistic interpretation developed below, poorly matched to the behavior's neural organization, whatever category they belong to. That

interpretation is provisional, and Section 4.1 sets out the evidence supporting it together with the limits on what the evidence can carry. The point applies symmetrically, since a reinforcement-based counterconditioning protocol and a graded-exposure desensitization protocol are expected to have limited effect for the same mechanistic reasons, and the framework's probability variable should reflect that, setting aside the practitioner's confidence in the methodological category.

4.1 The Neural Organization of Predation

Broseghini et al. (2024) coded sixty video recordings of dogs performing predatory acts, with two observers working independently, and resolved the canine sequence into four functional phases, search, approach, chase and bite, each described at three hierarchical levels and varying by working-dog group. That ethogram is the observational base assumed here for dogs. The longer sequence familiar from applied writing, running from orienting through dissection and consumption, comes from a general-readership account of canine evolution (Coppinger and Coppinger, 2001) and has no comparable coding base.

Circuit-level work on hunting is almost entirely rodent. Prey detection reaches the superior colliculus, from which an excitatory projection to the zona incerta triggers the behavior (Shang et al., 2019). Photoactivation of GABAergic zona incerta neurons promotes hunting of live and artificial prey while photoinhibition impairs it, a single 100-millisecond pulse suffices to trigger bite attacks on bedding material, and the same neurons discriminate prey from non-prey targets (Zhao et al., 2019). GABAergic projections from the lateral hypothalamus to the ventral periaqueductal gray drive predatory attack, and the central nucleus of the amygdala sends mainly inhibitory projections to the same region, where activation enhances pursuit (Franklin, 2019; Li et al., 2018). Periaqueductal activity follows a sequential pattern time-locked to distinct hunting actions, and disrupting inputs from the amygdala, hypothalamus, and zona incerta impairs both the sequence and hunting success (Yu et al., 2021). Han et al. (2017) separated pursuit from kill-bite in the central nucleus of the amygdala, a projection to the periaqueductal gray mediating pursuit and one to the brainstem reticular formation coordinating the musculature of the lethal bite, with targeted lesions disrupting each separately.

Dopaminergic involvement has been traced along two routes serving different phases. Huang et al. (2021) found a glutamatergic projection from the superior colliculus onto dopamine neurons of the substantia nigra pars compacta, whose activation triggers dopamine release in the dorsal striatum and increases the speed and frequency of approach during hunting, while inactivation impairs appetitive locomotion. Wang et al. (2025) traced a

dopaminergic projection from the ventral tegmental area to the medial olfactory tubercle that mediates an odor-guided preference between prey items, with chemogenetic inhibition of the pathway abolishing the preference and activation restoring it. Reviewing this literature, Zhao et al. (2023) describe zona incerta neurons as intrinsically rewarding and as energizing hunting.

That evidence establishes that the predatory drive is appetitive, that activation of the circuits generating it is itself rewarding, and that the drive is separable from feeding, since collicular pathway activation provokes attack on prey in satiated as well as hungry animals and its inactivation leaves daily food intake unchanged (Shang et al., 2019). It does not establish that performance of each element of the sequence reinforces the element preceding it, which is a separate proposition that no study has tested. The account of the sequence as intrinsically reinforcing is treated here as a mechanistic reading of the appetitive evidence, recorded as such on the assessment sheet, with the status of a demonstrated learning mechanism left open.

Counterconditioning and systematic desensitization act on a different learning problem. Both revise what a stimulus predicts, and the neural work locates that revision in basolateral amygdala plasticity under infralimbic prefrontal control, with central-nucleus output governing response expression (LeDoux, 2000; Quirk and Mueller, 2008).

The framework needs to distinguish between pathways. The central nucleus of the amygdala is limbic and is central to hunting, so no argument that predation bypasses the limbic system can be sustained. The amygdalar contribution to hunting runs through central-nucleus output to periaqueductal motor control, while the contribution of counterconditioning revises runs through the basolateral route by which the threat value of a stimulus is learned and relearned. A protocol that changes what a stimulus predicts about threat acts on a different input to a partly shared output, and the prey recognition gating the predatory sequence is not a threat appraisal that safety learning revises.

Appraisal is not absent from predation, and the framework does not claim that it is. Prey and non-prey discrimination precedes the sequence (Zhao et al., 2019), and defensive state competes with hunting instead of standing outside it. Rossier et al. (2021) recorded periaqueductal neurons encoding both approach and defensive behaviors, active selectively when animals showed heightened fear, and reported that lateral hypothalamic GABAergic neurons promote predation partly by suppressing defensive responses. Older feline literature points in the same direction, with amygdaloid stimulation modulating hypothalamically elicited attack (Block et al., 1980). Predation is therefore open to modulation through defensive competition, which is the

mechanism by which a maintained aversive contingency can suppress the expressed sequence without altering the motivation organizing it.

Every circuit study cited here used mice or rats. No comparable work exists for any canid, so applying this account to canine chase extrapolates as freely as the discount parameters conceded in Section 2.2 do, and the feline studies the applied literature leans on concern quiet biting attack elicited by hypothalamic stimulation, which is a different behavior. There is also a plain gap: appetitive counterconditioning has never been tested against predatory pursuit in any species. Absence of a test is not evidence of failure. Section 4.1 is therefore a provisional mechanistic interpretation, and the probability estimates it informs in Section 4.2 carry an epistemic flag on that basis.

4.2 Implications for the Framework

The implications for the framework are threefold. First, the probability that a conventional therapeutic procedure will modify predatory motivation should be set low, whatever category the procedure is drawn from, because the mechanistic interpretation in Section 4.1 predicts that the procedure acts on a route other than the one organizing the behavior. A counterconditioning protocol using high-value reinforcement and a desensitization protocol using gradually intensified aversive stimulation have, on that reading, the same low probability of modifying the motivation through the mechanism each assumes. The estimate is a prediction from mechanism. No study in any species has tested appetitive counterconditioning against predatory pursuit, so any application resting on it raises an epistemic flag on the probability variable.

The mechanistic claim concerns the route a procedure assumes. Other routes remain open, since suppression under a maintained contingency acts on expression instead of on the motivation organizing it, through the defensive competition described in Section 4.1. Johnson and Wynne (2024) reported that the six dogs completing an electronic-collar condition stopped pursuing a mechanical lure within one or two sessions, that none of the eleven in two reinforcement-based conditions stopped in training or in any test, and that four of the six withheld pursuit once the arena and lure were changed. Two dogs left that condition at the twenty-shock ceiling their ethics committee permitted, and every dog in it vocalized on stimulation, significantly more than in the reinforcement-based conditions, $H(2) = 7.14$, $p = 0.03$, which the authors read as evidence of pain. That paper's abstract states instead that 67 percent chased the lure in that test. The results section is followed here, and on the abstract's reading, the Section 6.2 ratio falls further, strengthening the conclusion reached there. Christiansen et al. (2001) tested 114 hunting dogs against domestic sheep under a contingency delivering one second of stimulation at 3000 V and 0.4 A on approach to the flock; 17 of them received shocks in the first year, and at the one-year retest attack frequency had fallen

from 42.1 to 11.4 percent. A procedure reliably eliciting pain-related vocalization in every treated individual cannot be scored at the low end of the intensity scale, whatever its efficacy, and that ceiling is Threshold 1 operating inside a published protocol.

Contingency control also operates at intensities far below the collar studies, and on a reinforcer taken from the predatory domain itself. Marschark and Baenninger (2002) followed seven naive herding dogs and their inexperienced owners across several months of weekly sessions, coding every trainer vocalization as blocking, which withholds access to sheep, or access, which grants it. Blocking ran consistently above access, and mean blocking fell from 7.83 to 6.87 commands per minute between the first and last five sessions ($n = 7$, $T+ = 28$, $p = 0.008$) while access rose from 4.04 to 4.57 without reaching significance ($p = 0.188$). The authors conclude that procedures beyond positive reinforcement may be necessary where instinctive motor patterns are at issue. Neither food nor electronic stimulation featured in the protocol. Their account of why the contingency works carries the weight here: without the dog's instinctive orienting and approach to livestock, access to sheep would have no value as a reinforcer and its withdrawal none as a punisher. Herding brings the predatory sequence under contingency control while leaving the motivation intact, the kill response bred out, and gripping retained, which is the distinction between suppression of expression and modification of motivation drawn above.

Second, the alternative-exhaustion threshold operates differently for predation cases, since conventional therapeutic procedures of any category fall outside the relevant alternatives. These include contingency control, environmental management to prevent rehearsal of the predatory sequence, channeled outlets for components of the sequence in safe contexts, and management of the wider environment to reduce exposure to predatory stimuli. Prevention of rehearsal has its own mechanistic rationale. Hunting competence is acquired and refined with practice, naive mice taking 391 ± 109 seconds to initiate hunting against 43 ± 11 seconds once trained over five to nine days, with defensive responses to prey falling across the same period (Rossier et al., 2021), so exposure to a prey-like target is not a neutral manipulation and a graded-exposure protocol in this domain risks strengthening what it aims to reduce. A procedure proposed for predation must document the exhaustion of these alternatives.

Third, as already mentioned, the framework treats predation as a separate analytical category that requires species-typical understanding of the predatory motor sequence and breed-specific selective history. Some breeds have been selected for amplified expression of specific phases of the predatory sequence, with herding breeds expressing exaggerated stalk-chase phases, scent hounds expressing exaggerated trailing phases, and terriers expressing

the kill-bite phase under conditions where the predatory sequence is activated by small-prey stimuli. The framework's evaluation of any predation case must situate the individual within the relevant breed-typical predatory profile and within the species-typical understanding of which phases of the sequence are most resistant to modification.

4.3 Predation and the Term "Predatory Aggression"

The term "predatory aggression," common in clinical and applied literature, groups together two motivational domains that the present analysis, following comparative neurobiology, treats as distinct. Aggressive behavior, in the ethological sense, is social behavior regulating conflict over resources and social position, and the neurobiological literature treats it as an emergent property of a social behavior network responsive to the appraisal of threat or competition, an account set out in a review of the neural mechanisms of aggression in which predation does not figure (Nelson and Trainor, 2007). Predation involves feeding systems organized through subcortical motor sequences with dopaminergic support. The dissociation is supported by the separation of pursuit from kill-bite within distinct amygdalar projections (Han et al., 2017), by the separate circuits for predation and evasion in the hypothalamus (Li et al., 2018), and by the finding that activating the collicular hunting pathway provokes predatory attack on prey while leaving aggression toward conspecifics unaffected (Shang et al., 2019). This division is a working analytic one, open to revision as circuit evidence develops, and it carries practical weight, since grouping the two domains can lead to intervention proposals aimed at a mismatched substrate.

5. Operationalization

5.1 The Assessment Sheet

Practical deployment requires a standardized assessment sheet completed before any procedure, with a stated justification for long-term welfare or survival. Table 2 sets out the full sheet, following the order in Section 3.4 and the rationale given in the note beneath it. The duration convention adopted is recorded among the opening entries, so that sheets scored under different conventions are not compared. The sheet closes with the epistemic flags raised, the alternatives compared with their ratios, and a written welfare rationale an external reviewer could audit without the original assessor present.

Table 2: The assessment sheet completed for a single individual and a single proposed procedure

Field	Entry
Individual and context	
Behavior and motivational domain	
Less invasive alternatives already tried	

Field	Entry
Duration unit convention adopted	
Intensity (I), 1 to 5	
Duration (D)	
Frequency (F)	
Reversibility (R), 0 to 1	
Short-term cost, $C = I \times D \times F \times (2 - R)$	
Magnitude (M), 1 to 5	
Probability (P), 0 to 1	
Expected duration of outcome (T)	
Outcome, $B = M \times P \times T$	
Delay (t) and discount rate, with source	
Present value, $PV(B)$	
Net welfare value, $NWV = PV(B) - C$	
Benefit-cost ratio, $PV(B) / C$	
Threshold 1, intensity below 5	
Threshold 2, reversibility at or above 0.4	
Threshold 3, alternatives documented and exhausted	
Epistemic flags raised, on P, on the discount rate, on Threshold 3 evidence	
Alternatives scored for comparison, with their ratios	
Written welfare rationale	

Note: Rows are completed in the order shown. The first four precede any scoring, so that the evidence bearing on Threshold 3 is gathered independently of the result, and the three threshold verdicts follow the arithmetic, since alternative exhaustion cannot be adjudicated until the alternatives have themselves been scored and ranked.

A third party reading a completed sheet should be able to see where the assessor's judgment was strong, where it rested on weaker evidence, and where alternative plausible scorings would have produced different verdicts.

5.2 Species-Specific Rubrics

The intensity and reversibility scales require species-specific rubrics that no responsible application can generalize across taxa. For companion dogs, the rubric should use cognitive-bias measures, which Mendl et al. (2009) support at face value while naming feeding motivation, activity, and learning as alternative explanations, and the critical evaluation of physiological welfare indicators in Cobb et al. (2016), a caution applying equally to studies critical of a training methodology and to studies defending it. Cobb et al. (2025) review heart rate variability and other physiological indicators as alternatives, noting that the relationship between behavioral and heart-rate-variability measures remains contested, and the framework adopts their cautions for any physiological scoring within the cost function.

For horses, the ridden-horse pain ethogram provides a validated starting point (Dyson et al., 2018), with the equitation-science welfare literature supplying the normative framing instead of an instrument (McLean

and McGreevy, 2010). For captive wild individuals prepared for translocation, rubrics can build on the conservation-behavior literature on anti-predator training (Greggor et al., 2020; Moseby et al., 2012; Shier and Owings, 2006) and, for captive settings, on the judgment-bias design guidance in Bethell (2015).

A full operational version would include a library of validated species-specific rubrics maintained as a living resource. The framework should not be deployed where the indicator base is too thin for reliable scoring, and in such cases the assessor defaults to the categorical thresholds, records an epistemic flag, and leaves unscored those variables the evidence cannot support.

5.3 The Judgment Bias Test

The cognitive judgment bias test, extended to dogs, pigs, sheep, macaques and starlings, supplies a method for assessing medium-term affective state that can serve as external validation of framework predictions, described in its own literature as promising instead of established (Bethell, 2015; Mendl et al., 2009). Where a procedure is scored as producing net positive welfare after a specified recovery period, judgment-bias testing at that point should show no more pessimistic response than baseline, and ideally a more optimistic one. Systematic use of the test as a post hoc validation procedure would provide an empirical basis for refining the scoring rubrics and discount-rate calibrations.

The same logic applies to physiological markers of chronic stress. Bateson and Poirier (2019) evaluate two, leukocyte telomere length and hippocampal volume, as candidate biomarkers of cumulative lifetime experience, while cortisol measures carry the limitations Cobb et al. (2016) record. Less sensitive than behavioral tests to transient affective states, these measures are valuable for identifying procedures whose cumulative welfare cost was underestimated beforehand. Where post hoc validation diverges from pre-intervention scoring, the assessment record must be revised.

6. Worked Examples

The framework is illustrated below with three cases from applied ethology, conservation biology, and everyday companion-animal training, each presented as a completed assessment sheet in Table 3. The cases are stylized for exposition, and every numerical input in this section is an illustrative value chosen to show the framework in operation. No figure in Section 6 should be read as an empirical estimate, the two exceptions being the counts reported by Johnson and Wynne (2024) and Christiansen et al. (2001), identified as such where they appear. A successful application produces a completed and auditable sheet carrying a benefit-cost ratio, a verdict on each of the three categorical thresholds, a record of every epistemic

flag raised, and a ranking against the alternatives scored under the same convention. That record is the intended outcome, whatever verdict it carries, and all three cases achieve it.

6.1 Anti-Predator Conditioning

Consider a captive-bred population of an endangered marsupial being prepared for release into a habitat populated by introduced predators. The published literature is mixed. Controlled exposure to predator models, paired with alarm calls and mild aversive stimuli, can strengthen anti-predator responses, subject to the constraints Griffin et al. (2000) set out, among them whether isolation from predators is ontogenetic or evolutionary. In situ exposure to a low density of feral cats over 18 months lengthened flight-initiation distances in reintroduced burrowing bettongs, on behavioral measures with no survival outcome (West et al., 2018). Whether such gains convert into post-release survival is less settled. Moseby et al. (2012) recorded markedly stronger avoidance in trained greater bilbies, with nine of ten trained and eight of ten untrained individuals surviving six months after release, and the population extinct at nineteen. Shier and Owings (2006) reported higher one-year survival in captive-born juvenile black-tailed prairie dogs trained against live predators. Tetzlaff et al. (2019) pooled 108 effects from 41 studies across four vertebrate classes and reported survival odds of 2.14 for anti-predator training (95 percent CI 1.34 to 3.40). The welfare calculus in these programs is rarely made explicit, trading short-term distress during training for a reduced probability of predation after release, on an evidence base whose strength varies by taxon and release context.

Applying the framework, the intensity of each training episode might score $I = 2$, corresponding to a brief, moderate alarm, consistent with the species-typical response to predator signals. Duration might be $D = 60$ seconds per episode. The frequency might be $F = 6$ over a two-week conditioning period. Reversibility scores $R = 0.8$, reflecting rapid behavioral recovery and the absence of generalized fear conditioning outside the training context. The computed cost is $C = 2 \times 60 \times 6 \times (2 - 0.8) = 864$ cost units.

The outcome calculation begins with magnitude $M = 4$, reflecting the welfare implications of continued survival in a species-typical environment. Probability, taken from the 2.14 survival odds Tetzlaff et al. (2019) report for anti-predator training and translated into a proportional probability, scores an illustrative $P = 0.6$. The translation from a relative survival improvement to an absolute probability is a modeling choice recorded on the sheet, so that a reviewer can substitute another and see its effect, and the divergence between the behavioral and survival literatures noted above raises an epistemic flag on the variable. Expected duration is $T = 730$ days, approximating two years of post-release survival during which the learned response remains functional,

bounded as Section 3.3 requires by the actuarial expectation for the release context, which in Moseby et al. (2012) proved shorter. The raw outcome is $B = 4 \times 0.6 \times 730 = 1,752$ units.

The discount function, using a hyperbolic parameter $\delta = 0.001$ per day appropriate for a small mammal based on interpolation from rat and kangaroo rat data, and delay $t = 14$ days between intervention and the beginning of outcome realization, gives $PV(B) = 1,752 / (1 + 0.001 \times 14) \sim 1,728$ units. The Net Welfare Value is thus $1,728 - 864 = 864$ units, and the benefit-cost ratio is 2.00.

The three thresholds are then tested. Threshold 1 is satisfied at $I = 2$ and Threshold 2 at $R = 0.8$, both well clear of their limits. Threshold 3 requires that the alternative-exhaustion documentation credibly assert that less aversive methods have failed to produce comparable post-release survival in similar programs. On the published record, the requirement is met for the behavioral outcome, and the documentation is entered on the sheet with the survival caveat recorded above.

Two epistemic flags are raised. The discount rate is interpolated from related species, since direct delay-discounting data for the focal species are unavailable, and the probability estimate rests on a meta-analytic pooling across programs whose individual results diverge. Both flags are informational and leave the procedure eligible. Sensitivity analysis varying δ across plausible values for small mammals leaves the verdict unchanged. The framework's verdict is that the procedure is welfare-defensible on the behavioral outcome, qualified by both flags, each documented in the assessment sheet.

6.2 Recall Training under Predation Risk

Consider a working collie living on a sheep farm who has twice broken recall to chase livestock. The practitioner is considering three interventions: supplementary positive-reinforcement recall training with high-value food reinforcement, supplementary recall training using aversive stimulation, and environmental management restricting the dog's off-lead access. All three are scored against a single outcome, preventing livestock-chase events during the dog's remaining life, with the magnitude term common, so comparison turns on probability, cost, and delay. The first analytical step identifies the motivational domain. The behavior as described is predatory in character, and the mechanistic considerations developed in Section 4 apply directly, subject to the limits recorded there.

For the supplementary positive-reinforcement recall training, the framework calculates as follows. Cost is low, with $I = 1$, $D = 60$ seconds of welfare-relevant exposure per episode, $F = 30$ episodes, $R = 1$, for $C = 1,800$ cost units. The duration records only the welfare-relevant portion of each

episode, and the same convention applies to every option compared below. Magnitude is high, $M = 5$. The probability of modifying the predatory behavior through this procedure is low on the mechanistic reading in Section 4. An illustrative $P = 0.1$ reflects the residual probability that the protocol produces some recall improvement under non-predatory conditions while leaving the predatory behavior unchanged, and an epistemic flag is raised because no direct efficacy data exist for this intervention in this domain. $T = 2,555$ days. Raw outcome $B = 5 \times 0.1 \times 2,555 = 1,278$ units. Discounted using an illustrative $\delta = 0.001$ per day and delay $t = 10$ days, $PV(B) \sim 1,265$ units. $NWV = 1,265 - 1,800 = -535$ units, and the benefit-cost ratio is 0.70. Section 3.1 forbids reading the sign of the Net Welfare Value as a verdict, so the ratio carries the analysis, and 0.70 is the highest of the three supplementary training scenarios and still an order of magnitude below the management option scored later.

For the supplementary aversive-stimulation training, the framework calculates as follows. Intensity, depending on the equipment and protocol, might score $I = 2$ for a brief startle paired with the recall signal or $I = 4$ for sustained electronic-collar stimulation, and the two scenarios diverge sharply at that point. The $I = 2$ scenario, scored on the same convention, produces $C = 2 \times 60 \times 30 \times (2 - 0.85) = 4,140$ cost units. Probability turns on which outcome is claimed and on which protocol produced the evidence for it, and the evidence for suppression came from electronic collars at working intensity, as Section 4.2 records. No efficacy data exist for the $I = 2$ protocol in a predatory-chase context, so its probability is extrapolated downward from the collar studies, and an illustrative $P = 0.15$ is entered under an epistemic flag, above the reinforcement-based estimate because a contingency is present and well below the collar estimate because the stimulus is not comparable. $T = 2,555$ days. Raw outcome $B = 5 \times 0.15 \times 2,555 = 1,916$ units, discounted at the same rate and delay to $PV(B) \sim 1,897$ units. $NWV = 1,897 - 4,140 = -2,243$ units, and the benefit-cost ratio is 0.46. Threshold 1 is satisfied at $I = 2$ and Threshold 2 at $R = 0.85$, so the procedure clears the categorical gates and ranks below every other option in the case. A second caution belongs on the sheet, since the cost is scored over a bounded intervention period of thirty episodes, while a suppression outcome sustained by a maintained contingency implies continued exposure across the individual's life; an assessor scoring frequency over that horizon returns a higher cost and a lower ratio still.

The $I = 4$ scenario carries the efficacy evidence and the intensity that produced it. Four of the eight dogs allocated to the electronic-collar condition in Johnson and Wynne (2024) reached suppression that transferred out of the training context, as Section 4.2 records. Christiansen et al. (2001) report a larger reduction persisting at one year, on a before-after design without a control for repeated exposure or maturation, so the two sources bracket the

estimate and an illustrative $P = 0.5$ sits at its conservative end. Raw outcome $B = 5 \times 0.5 \times 2,555 = 6,388$ units, discounted to $PV(B) \sim 6,325$ units. The cost is $C = 4 \times 60 \times 30 \times (2 - 0.6) = 10,080$ units, giving a net welfare value of -3,755 units and a benefit-cost ratio of 0.63. Threshold 1 is satisfied at $I = 4$ but with the most demanding documentation requirements, and Threshold 2 is satisfied at $R = 0.6$ only marginally. The combination of high intensity and marginal reversibility substantially increases the documentary burden. An epistemic flag is raised on this probability as well, since it rests on eight animals in an experimental setting with no follow-up testing of durability, and a flag on the better-evidenced estimate is as necessary as one on the weaker.

Entering $P = 0.5$ against the $I = 2$ cost returns a ratio of 1.53 and makes the milder option appear defensible. That pairing is not available, because the 0.5 rests on two protocols scored at $I = 4$, one in which every dog exposed vocalized on stimulation and two left the condition at an institutional ceiling, the other delivering 3000 V and 0.4 A. Efficacy and intensity belong to one protocol and are scored together. Read that way, no aversive-stimulation scenario in this case returns a ratio above one, and the strongest stands at 0.63 against 0.70 for the reinforcement-based option, with both far below management at 12.05 under the exposure convention adopted here. D records the welfare-relevant portion of an episode in place of stimulus-delivery duration, one second in Christiansen et al. (2001), and is held constant with F across the three training scenarios so the comparison isolates intensity and reversibility. That convention cuts against the aversive options, and the paragraph on conventions below reports what happens under the alternatives.

For the environmental management approach, the framework calculates as follows. The intervention modifies the dog's off-lead access through fencing, supervised exercise, and a structured outlet for components of the predatory motor sequence in safe contexts. $I = 1$ for the brief restraint involved in transitions, $D = 30$ seconds of welfare-relevant exposure per episode, $F = 30$ episodes, $R = 1$, for $C = 900$ cost units. This arrangement persists across the outcome horizon, so the sheet must record why only the transitions are scored: a maintained aversive contingency generates an episode at each application, whereas management substitutes species-typical outlets, so its steady state carries no recurring cost. An assessor judging otherwise, on confinement or restricted agency, scores frequency across the full horizon, which returns a cost of roughly 76,650 units and a ratio of 0.14. The judgment is not idle, since the least-inhibitive commitment of the LIFE model treats continuous restriction of choice as a welfare cost in its own right (Fernandez, 2024), and Section 7 records that the cost function has no term for it. Magnitude is high at $M = 5$; the illustrative probability of preventing livestock-chase events is high at $P = 0.85$ because the intervention removes access to the eliciting stimulus while leaving predatory motivation unaddressed; and the

expected duration is $T = 2,555$ days. No published efficacy study of this management package in a livestock-chase context exists either, so the first condition of Section 3.5 applies here as it does to the training options, and the estimate carries an epistemic flag. Physical prevention of access is a more reliable route to the outcome than behavior change, which is why the value is set high, and a fence that is climbed or a supervision lapse moves it down. Raw outcome $B = 10,859$ units, discounted over a one-day delay to $PV(B) \sim 10,848$ units, giving $NWV = 9,948$ units and a benefit-cost ratio of 12.05, the highest of the four scenarios compared.

The comparison settles the case at Threshold 3 before the arithmetic is consulted. Environmental management is a documented, less invasive alternative returning 12.05, against 0.70 for the positive-reinforcement option and 0.63 and 0.46 for the two aversive-stimulation scenarios, and its availability, with supporting efficacy evidence, means no supplementary training scenario satisfies the alternative-exhaustion requirement. All three are refused there, whatever their computed values, and the refusal would stand had a training option returned the higher ratio, which is what a categorical gate is for. The ranking is also insensitive to the probability estimates. Setting probability to its ceiling in each training scenario returns maximum ratios of 7.03, 3.06, and 1.25 against 12.05 for management, so under the convention adopted here no admissible probability makes any training option outrank the management alternative, and the comparative result does not depend on the mechanistic reading in Section 4.1.

Intensity and magnitude are ordinal, so an ordering counts only where it survives the plausible alternatives. Under all joint perturbations of one scale point on intensity and magnitude and 0.1 on reversibility and probability, with scores clamped at their scale bounds and probability floored at 0.01, management returns a worst case of 3.87 against a best case of 1.67 for any training scenario, so its advantage is a dominance relation independent of the interval assumption. The three training scenarios overlap across the same set, at 0.03 to 1.41, 0.07 to 1.67 and 0.30 to 1.08, so their ordering is reported as unresolved, and the 0.30 bound is computed at $I = 5$, which Threshold 1 refuses, narrowing that scenario's admissible range further. Any application should carry the same check.

Conventions need the same treatment as scores, since Section 3.1 admits only a convention that can be applied to every option in a case. Scoring duration at stimulus-delivery time, one second in Christiansen et al. (2001), would return ratios of 27.5 for the $I = 2$ scenario and 37.6 for the $I = 4$ scenario, placing both above every other option in the case. It is inadmissible here because a reinforcement program and a management arrangement deliver no discrete stimulus, so the convention cannot apply and the comparison would cease to be like-for-like. Scoring frequency over the period each intervention

generates welfare-relevant episodes does pass that test, since a bounded program keeps its episode count while a persistent arrangement takes the outcome horizon. On that convention, management costs 76,650 units for a ratio of 0.14; the maintained $I = 2$ and $I = 4$ contingencies cost 352,590 and 858,480 units for ratios of 0.005 and 0.007, and the bounded reinforcement program is unchanged at 0.70. The two admissible conventions disagree about which of the two non-aversive options ranks first, and agree that both aversive scenarios rank last. Neither disturbs the Threshold 3 refusal, which turns on the availability of a less invasive documented alternative instead of on any ratio.

The case supports three conclusions. First, it discriminates among options on mechanism and evidence instead of on methodological label. The three training scenarios differ in the evidence available for the outcome each claims, and the one with the strongest support is still refused because the intensity that produced that evidence raises its cost beyond what the evidence can carry and because a less invasive alternative outranks it. Second, management scores well here because it addresses the case structurally and leaves the motivational domain unaltered. Third, the framework takes no position in the methodological debate over training methods in dogs, applying one analytical structure across categories. Since the comparison depends on scoring every option under one duration-and-frequency convention, any application should report that convention and confirm the ordering holds under plausible alternative exposure assumptions.

6.3 Cooperative Husbandry as a General Training Application

The two cases above involve survival outcomes and a recall problem with a predatory component. A third case shows this framework in everyday work: a dog shows escalating distress during nail trimming, making routine care difficult and leading the owner to avoid it. The practitioner suggests systematic desensitization with positive reinforcement, training the dog to accept paw handling and clipping in steps. The behavior to be changed is the defensive response to restraint and handling, and the motivational identification is mild defensive arousal, a profile distinct from predatory or social motivation, which sets the expectation that a graded counterconditioning approach matched to that system carries a high probability of success.

The short-term cost is low. Intensity is set at 1, reflecting the mild and transient arousal of a well-paced desensitization session; duration at 120 seconds of mild exposure per session; frequency at 20 sessions across the program; and reversibility at 1, since the procedure leaves no lasting harm and can be halted at any point. The cost function gives $C = 1 \times 120 \times 20 \times (2 - 1) = 2,400$ cost units.

The long-term outcome is substantial. Magnitude is set at 4 because reliable acceptance of husbandry removes a recurring source of distress and allows veterinary and grooming care to proceed without force throughout the individual's life. Probability is set at 0.85, reflecting strong support for graded counterconditioning of handling in the applied literature, and the duration is set at 3,650 days (approximately 10 years). The raw benefit is $B = 4 \times 0.85 \times 3,650 = 12,410$ units. With a 30-day delay to a settled outcome and a discount rate of 0.001 per day, the present value is $12,410 / (1 + 0.001 \times 30) \sim 12,049$ units.

The Net Welfare Value is $12,049 - 2,400 = 9,649$ units, and the benefit-cost ratio is 5.02. Intensity falls far below any plausible ceiling and reversibility sits at the most permissive floor value, so Thresholds 1 and 2 are satisfied without difficulty. Threshold 3 requires comment, since the alternatives on the sheet, forced restraint and sedation for routine care, are more invasive than the procedure proposed. Where a procedure already occupies the least invasive end of the available range, the threshold is satisfied trivially, and the sheet should record that instead of implying a less invasive option was tried and found wanting. No epistemic flag is raised, the technique being well documented. The same structure that scrutinizes costly or doubtful procedures also records why low-cost, well-evidenced ones are justified.

Table 3: The three worked cases, with inputs, calculations, threshold verdicts, and epistemic flags recorded for each

Assessment step	Case 1. Anti-predator conditioning	Case 2. Recall under predation risk	Case 3. Cooperative husbandry
1. Individual and behavior	Captive-bred marsupial, predator-naïve translocation candidate	Working collie, breaking recall to chase livestock	Companion dog, escalating distress at nail trimming
Motivational domain	Defensive, anti-predator	Predatory	Mild defensive arousal
Alternatives recorded before scoring	Simple habituation, unsuccessful in comparable programs, with survival evidence mixed across taxa	Three supplementary training scenarios, all scored	Forced restraint or sedation for routine care, both more invasive than the procedure proposed, so Threshold 3 is satisfied trivially
Duration convention	Seconds per episode, days for outcome	Seconds per episode, days for outcome	Seconds per episode, days for outcome
2. Cost inputs, I / D / F / R	2 / 60 / 6 / 0.8	1 / 30 / 30 / 1	1 / 120 / 20 / 1
Short-term cost, C	864	900	2,400
3. Outcome inputs, M / P / T	4 / 0.6 / 730	5 / 0.85 / 2,555	4 / 0.85 / 3,650
Raw outcome, B	1,752	10,859	12,410
4. Delay and discount rate	14 days, 0.001 per day, interpolated	1 day, 0.001 per day, illustrative	30 days, 0.001 per day, illustrative
Present value, PV(B)	1,728	10,848	12,049
5. Net welfare value	864	9,948	9,649

Assessment step	Case 1. Anti-predator conditioning	Case 2. Recall under predation risk	Case 3. Cooperative husbandry
Benefit-cost ratio	2.00	12.05	5.02
6. Threshold 1, intensity below 5	Satisfied, I = 2	Satisfied, I = 1	Satisfied, I = 1
Threshold 2, reversibility at or above 0.4	Satisfied, R = 0.8	Satisfied, R = 1	Satisfied, R = 1
Threshold 3, alternatives exhausted	Satisfied on the published record	Satisfied, and refuses all three training scenarios	Satisfied, restraint and sedation carry higher cost
7. Epistemic flags	Two, on the discount rate and on the probability estimate	One on each option, on the probability estimate	None
Intended outcome, a complete and auditable sheet	Achieved	Achieved	Achieved
Verdict recorded	Welfare-defensible on the behavioral outcome, qualified by both flags	Environmental management selected, all three training scenarios refused at Threshold 3	Welfare-defensible, no qualification

Note: Every option in Case 2 carries a flag on its probability estimate. The three supplementary training scenarios returned ratios of 0.70, 0.63 and 0.46 and were refused at Threshold 3, a documented and less invasive alternative being available. All values are illustrative and are scored under a common duration convention, cost duration in seconds and outcome duration in days.

7. Limitations

Any framework that reduces complex welfare judgments to numerical outputs carries risks that deserve explicit acknowledgment. Eleven are addressed here, so that the framework is deployed with the epistemic humility its design demands.

The first concern is that the numerical precision of the Net Welfare Value equation is illusory given the measurement uncertainty involved, a point developed at the head of Section 3 and governing everything that follows. Practitioners should report sensitivity analyses showing how the ratio changes under plausible alternative scorings, treating the surviving ranking and the approximate magnitude as the informative outputs (Beausoleil and Mellor, 2015; Browning, 2022).

The second concern is that the probability variable is often the weakest parameter, as efficacy data for specific interventions in specific species and contexts are scarce. The epistemic flag mitigates this by separating evidentially thin assessments from strong ones without disqualifying the former, which lessens the problem while leaving it in place. In novel contexts, the framework should sit beside practitioner judgment and inform it, its outputs treated as one input to a wider deliberative process.

The third concern is that the discount rate is species-specific and individual-specific, and the available literature supplies only rough interpolations for most species. The range and construct problems are set out in Section 2.2. Research priorities therefore include mapping delay-discounting parameters across a broader taxonomic range and across longer horizons, with particular attention to species commonly involved in training and conservation contexts.

The fourth concern is that the framework addresses individual welfare and leaves population- and conservation-level considerations to other tools. Genetic diversity, population viability and ecosystem function cannot be reduced to the individual-welfare calculus presented here, and where they are at stake the framework should sit within a wider conservation-ethics structure as one decision input among several (Beausoleil et al., 2018).

The fifth concern is that the framework rests on a utilitarian core bounded by categorical thresholds, and this hybrid structure will draw objection from those holding strong deontological positions on the permissibility of any deliberate welfare-relevant intervention. The structure reflects the reasoning thoughtful practitioners use, and defending it as a meta-ethical position lies beyond the scope of this paper. A deontologist applying the framework would set the intensity ceiling lower, the reversibility floor higher, and the alternative-exhaustion requirement more stringent, and the framework accommodates all three without structural modification. That accommodation carries a consequence worth stating: the framework displays the commitments an assessor brings to it and does not adjudicate between them. Two people who disagree can locate the parameter on which they disagree, which is a real contribution and a modest one.

The sixth concern is that the quantitative apparatus could be misapplied in either direction, lending unwarranted legitimacy to procedures that should not be performed or refusing procedures with substantial welfare benefit. The risk runs symmetrically, and three structural features respond to it. The categorical thresholds hold firm, so favorable arithmetic alone leaves them in force. The epistemic flag makes the framework's confidence state explicit at the point of decision. The assessment sheet generates an auditable record that survives the assessor. None of this prevents misuse by a determined bad-faith actor, and the framework makes no such claim, but each raises the documentary cost of misuse and exposes it to later review.

A seventh consideration concerns the framework's relationship to the wider debates in animal training. It is method-neutral by design, importing none of the asymmetries common to position statements in this area. Ziv (2017) reviewed seventeen studies of aversive methods in dogs and reported an association with stress- and fear-related responses in a narrative review combining owner surveys, direct observation, controlled experiments, and a

case report, with no formal appraisal of study quality and causality left open by the author. The case advanced on the other side rests on a correspondingly small efficacy literature, principally Christiansen et al. (2001), Salgirli et al. (2012) and Johnson and Wynne (2024), each limited in the ways Sections 2.4 and 4.2 record. Neither body of work supports a general verdict.

An eighth concern is that the functional forms are modeling choices open to empirical refinement, as Table 1 records and Section 3.6 argues. A research group preferring a different monotonic combination of the same parameters can substitute it, provided it states the substitution and applies it consistently. The cost function is episodic, with no term for continuous restriction of agency, so options that constrain behavior without discrete episodes are advantaged by the structure of C instead of by the evidence, a cost the least-inhibitive commitment of the LIFE model names and this framework undercounts. The property belongs to any cost function counted in episodes, so it bounds what method-neutrality can mean here. The framework is neutral in its evidentiary standards and its thresholds, and its cost function is not neutral in shape, which is why Section 6.2 tests conventions alongside scores.

A ninth concern, and the most consequential for the framework's ethics, is the legitimacy of discounting welfare across time. A long tradition in ethics holds that an individual's future welfare should not be devalued merely because it is distant. Section 2.2 sets out the framework's answer: the discount represents the individual's own temporal valuation, leaving the moral weight of distant welfare with the assessor, along with the concession that a choice rate is not the welfare experienced later. Where an assessor judges that a decision should give full undiscounted weight to the individual's future, the discount rate can be set to zero, and that choice is recorded as any other.

A tenth concern is that the outcome term combines magnitude, probability, and duration multiplicatively, which assumes a risk-neutral and roughly independent treatment of those dimensions. The product carries two commitments worth separating. Probability here is the assessor's stated credence with its basis recorded, since no reference class exists for a single individual in a single context, and the gain lies in a credence that can be challenged in place of one held silently. The product is accordingly bookkeeping for reasoning under uncertainty, and no ensemble exists across which a probability-weighted outcome is realized for one animal. An assessor with reason to treat a low-probability severe outcome differently from its expected value can record a risk adjustment beside the calculation.

An eleventh concern, and the one that most constrains how the output should be read, is that the cost and outcome terms are computed from durations measured on different physical scales, so both the magnitude of the Net Welfare Value and the position of its sign depend on the unit convention

adopted. Section 3.1 responds by taking the benefit-cost ratio as the decision-relevant statistic and by deferring its break-even value to the calibration described in Section 8. The framework therefore works comparatively, across the options a practitioner actually has, and cannot yet issue a verdict on a lone procedure.

8. Validation and Next Steps

The framework is offered as a first-order proposal that needs rigorous validation, and three directions take priority. The first is pilot validation against expert consensus, scoring documented cases and comparing outputs with independent welfare judgments from assessors with diverse methodological commitments. The second is empirical calibration of the intensity, reversibility, and discount-rate parameters for commonly used species, together with the break-even value of the benefit-cost ratio, the parameter that would let a lone procedure be judged instead of merely ranked against its alternatives. That item comes first, since until it is settled, the framework can order options without saying where defensibility begins. The third is inter-rater reliability testing, with the threshold verdicts as the priority, since they are the gates that determine the framework's decision behavior.

9. Recommendations for Practitioners

The framework earns its value in the hands of practitioners who understand the behavior they are trying to influence, since the quality of every parameter depends on knowledge of the individual. Ethology, for instance, allows the assessor to distinguish a defensive response from a predatory sequence, the distinction on which Section 4 rests. The science of behavior modification establishes what can be changed and how, and lets a practitioner read the training-methods literature critically. Affective and motivational neuroscience describes the systems that drive behavior, improving estimates of magnitude and probability. Welfare science supplies and qualifies the indicators that populate the cost and outcome terms. Ethics equips the practitioner to reason about the value questions the framework makes explicit and leaves open, the thresholds among them.

Conclusion

Applied ethology and animal training have accumulated sufficient conceptual and empirical resources to support a structured analytical tool for temporal welfare trade-offs, yet no such tool has been assembled from the available materials to date. The Temporal Welfare Trade-off Framework presented here is an attempt to close that gap by combining three mature research traditions, namely cumulative welfare assessment, temporal discounting from comparative psychology, and harm-benefit analysis

bounded by categorical thresholds and an epistemic flag. Each variable in the framework has an established measurement tradition, which is a claim about the provenance of its components. Whether their mathematical combination is valid is a separate question, since component validity does not transfer to the composite. The structure makes welfare arguments reviewable, and it records the absence of evidence as a parameter of the assessment in its own right.

The framework is method-neutral, applying the same analytical structure across methodological categories and letting the welfare-relevant variables settle each case. The training-methods literature is read critically, with the same evidentiary standards applied throughout. Predation is treated as a motivationally distinct domain in which conventional therapeutic procedures are, on a provisional mechanistic interpretation set out in Section 4.1, mismatched to predatory motivation, while suppression of the expressed sequence under a maintained contingency is a separate route carrying its own evidence and its own intensity cost. Every aversive-stimulation scenario scored in this paper was refused, at the alternative-exhaustion threshold and on comparative ranking, as was the reinforcement-based training option in the same case and on the same ground, and the framework reached those results from the published parameters of the studies most often cited in support of such procedures. The eleven limitations set out in Section 7 constrain present use. They run from the ordinal character of welfare scores and the weakness of the probability parameter to the range and construct problems in the discounting literature, the modeling choices embedded in the cost and outcome functions, and the commensurability of the two terms. Section 8 sets out the program that responds to them. Commensurability comes first, because until that ratio is calibrated, the framework can order options without saying where defensibility begins.

The framework should keep what it demonstrates now separate from what remains hypothetical. It demonstrates feasibility, internal transparency, auditability, and comparative operation, in that a completed sheet records the reasoning behind a welfare-relevant decision in a form another person can examine, and that it orders candidate procedures against one another. It does not demonstrate inter-rater reliability, parameter calibration, predictive validity against independent welfare outcomes, normative acceptability to the range of positions held in the field, or any effect on the quality of real welfare decisions. Each of those is an anticipated contribution awaiting the work described in Section 8, and the suggestion that a shared procedure could give practitioners and scientists with different commitments a common language for reasoning about the same case belongs in the second list, not the first. The framework is offered openly and without restriction, so that practitioners, ethics committees, and welfare auditors can pilot it, criticize it, adapt it to their own species and settings, and set it aside where better tools become available.

In return, it asks the practitioner to accept the responsibility the method is meant to support: to make a welfare-relevant choice on stated grounds that another person can examine.

Conflict of Interest: The author reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The author did not obtain any funding for this research.

References:

1. Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge University Press.
2. Bateson, M. (2016). Cumulative stress in research animals: Telomere attrition as a biomarker in a welfare context? *BioEssays*, 38(2), 201-212. <https://doi.org/10.1002/bies.201500127>
3. Bateson, M., & Poirier, C. (2019). Can biomarkers of biological age be used to assess cumulative lifetime experience? *Animal Welfare*, 28(1), 41-56. <https://doi.org/10.7120/09627286.28.1.041>
4. Beauchamp, T. L., & DeGrazia, D. (2020). *Principles of animal research ethics*. Oxford University Press.
5. Beausoleil, N. J., & Mellor, D. J. (2015). Advantages and limitations of the Five Domains Model for assessing welfare impacts associated with vertebrate pest control. *New Zealand Veterinary Journal*, 63(1), 37-43. <https://doi.org/10.1080/00480169.2014.956832>
6. Beausoleil, N. J., Mellor, D. J., Baker, L., Baker, S. E., Bellio, M., Clarke, A. S., Dale, A., Garlick, S., Jones, B., Harvey, A., Pitcher, B. J., Sherwen, S., Stockin, K. A., & Zito, S. (2018). "Feelings and fitness" not "feelings or fitness": The *raison d'être* of conservation welfare, which aligns conservation and animal welfare objectives. *Frontiers in Veterinary Science*, 5, Article 296. <https://doi.org/10.3389/fvets.2018.00296>
7. Bethell, E. J. (2015). A "how-to" guide for designing judgment bias studies to assess captive animal welfare. *Journal of Applied Animal Welfare Science*, 18(Suppl. 1), S18-S42. <https://doi.org/10.1080/10888705.2015.1075833>
8. Block, C. H., Siegel, A., & Edinger, H. (1980). Effects of amygdaloid stimulation upon trigeminal sensory fields of the lip that are established during hypothalamically elicited quiet biting attack in the cat. *Brain Research*, 197(1), 39-55. [https://doi.org/10.1016/0006-8993\(80\)90433-3](https://doi.org/10.1016/0006-8993(80)90433-3)

9. Brønstad, A., Newcomer, C. E., Decelle, T., Everitt, J. I., Guillen, J., & Laber, K. (2016). Current concepts of harm-benefit analysis of animal experiments: Report from the AALAS-FELASA Working Group on Harm-Benefit Analysis - Part 1. *Laboratory Animals*, 50(1, Suppl.), 1-20. <https://doi.org/10.1177/0023677216642398>
10. Broom, D. M. (1991). Animal welfare: Concepts and measurement. *Journal of Animal Science*, 69(10), 4167-4175. <https://doi.org/10.2527/1991.69104167x>
11. Broseghini, A., Lõoke, M., Guérineau, C., Marinelli, L., & Mongillo, P. (2024). Ethogram of the predatory sequence of dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, 279, Article 106402. <https://doi.org/10.1016/j.applanim.2024.106402>
12. Browning, H. (2022). Assessing measures of animal welfare. *Biology and Philosophy*, 37(4), Article 36. <https://doi.org/10.1007/s10539-022-09862-1>
13. China, L., Mills, D. S., & Cooper, J. J. (2020). Efficacy of dog training with and without remote electronic collars vs. a focus on positive reinforcement. *Frontiers in Veterinary Science*, 7, Article 508. <https://doi.org/10.3389/fvets.2020.00508>
14. Christiansen, F. O., Bakken, M., & Braastad, B. O. (2001). Behavioural changes and aversive conditioning in hunting dogs by the second-year confrontation with domestic sheep. *Applied Animal Behaviour Science*, 72(2), 131-143. [https://doi.org/10.1016/S0168-1591\(00\)00203-3](https://doi.org/10.1016/S0168-1591(00)00203-3)
15. Chung, S. H., & Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, 10(1), 67-74. <https://doi.org/10.1901/jeab.1967.10-67>
16. Cobb, M. L., Iskandarani, K., Chinchilli, V. M., & Dreschel, N. A. (2016). A systematic review and meta-analysis of salivary cortisol measurement in domestic canines. *Domestic Animal Endocrinology*, 57, 31-42. <https://doi.org/10.1016/j.domaniend.2016.04.003>
17. Cobb, M. L., Jiménez, A. G., & Dreschel, N. A. (2025). Beyond Cortisol! Physiological Indicators of Welfare for Dogs: Deficits, Misunderstandings and Opportunities. *Journal of Applied Animal Welfare Science*, 1-30. <https://doi.org/10.1080/10888705.2025.2572616>
18. Cooper, J. J., Cracknell, N., Hardiman, J., Wright, H., & Mills, D. (2014). The welfare consequences and efficacy of training pet dogs with remote electronic training collars in comparison to reward-based training. *PLOS ONE*, 9(9), Article e102722. <https://doi.org/10.1371/journal.pone.0102722>

19. Coppinger, R., & Coppinger, L. (2001). *Dogs: A startling new understanding of canine origin, behavior, and evolution*. Scribner.
20. Dyson, S., Berger, J., Ellis, A. D., & Mullard, J. (2018). Development of an ethogram for a pain scoring system in ridden horses and its application to determine the presence of musculoskeletal pain. *Journal of Veterinary Behavior*, 23, 47-57. <https://doi.org/10.1016/j.jveb.2017.10.008>
21. European Parliament and Council of the European Union. (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union*, L 276, 33-79.
22. Farm Animal Welfare Council. (2009). *Farm animal welfare in Great Britain: Past, present and future*. Farm Animal Welfare Council.
23. Fernandez, E. J. (2024). The least inhibitive, functionally effective (LIFE) model: A new framework for ethical animal training practices. *Journal of Veterinary Behavior*, 71, 63-68. <https://doi.org/10.1016/j.jveb.2023.12.001>
24. Franklin, T. B. (2019). Recent advancements surrounding the role of the periaqueductal gray in predators and prey. *Frontiers in Behavioral Neuroscience*, 13, Article 60. <https://doi.org/10.3389/fnbeh.2019.00060>
25. Fraser, D., Weary, D. M., Pajor, E. A., & Milligan, B. N. (1997). A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare*, 6(3), 187-205. <https://doi.org/10.1017/S0962728600019795>
26. Friedman, S. G. (2009). What's wrong with this picture? Effectiveness is not enough. *Good Bird Magazine*, 4(4), 12-18.
27. Gregg, T. R., & Siegel, A. (2001). Brain structures and neurotransmitters regulating aggression in cats: Implications for human aggression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 25(1), 91-140. [https://doi.org/10.1016/S0278-5846\(00\)00150-0](https://doi.org/10.1016/S0278-5846(00)00150-0)
28. Greggor, A. L., Berger-Tal, O., & Blumstein, D. T. (2020). The rules of attraction: The necessary role of animal cognition in explaining conservation failures and successes. *Annual Review of Ecology, Evolution, and Systematics*, 51, 483-503. <https://doi.org/10.1146/annurev-ecolsys-011720-103212>
29. Griffin, A. S., Blumstein, D. T., & Evans, C. S. (2000). Training captive-bred or translocated animals to avoid predators. *Conservation Biology*, 14(5), 1317-1326. <https://doi.org/10.1046/j.1523-1739.2000.99326.x>

30. Grimm, H., Olsson, I. A. S., & Sandøe, P. (2019). Harm-benefit analysis: What is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research. *Laboratory Animals*, 53(1), 17-27. <https://doi.org/10.1177/0023677218783004>
31. Han, W., Tellez, L. A., Rangel, M. J., Jr., Motta, S. C., Zhang, X., Perez, I. O., Canteras, N. S., Shammah-Lagnado, S. J., van den Pol, A. N., & de Araujo, I. E. (2017). Integrated control of predatory hunting by the central nucleus of the amygdala. *Cell*, 168(1-2), 311-324.e18. <https://doi.org/10.1016/j.cell.2016.12.027>
32. Honess, P., & Wolfensohn, S. (2010). The extended welfare assessment grid: A matrix for the assessment of welfare and cumulative suffering in experimental animals. *Alternatives to Laboratory Animals*, 38(3), 205-212. <https://doi.org/10.1177/026119291003800304>
33. Huang, M., Li, D., Cheng, X., Pei, Q., Xie, Z., Gu, H., Zhang, X., Chen, Z., Liu, A., Wang, Y., Sun, F., Li, Y., Zhang, J., He, M., Xie, Y., Zhang, F., Qi, X., Shang, C., & Cao, P. (2021). The tectonigral pathway regulates appetitive locomotion in predatory hunting in mice. *Nature Communications*, 12, Article 4409. <https://doi.org/10.1038/s41467-021-24696-3>
34. Hunter, T., van Rooy, D., McArthur, M., Bennett, S., Tuke, J., & Hazel, S. (2020). Mental health disease or preventable problem? Australian dog trainers' opinions about canine separation anxiety differ with training style. *Animals*, 10(8), Article 1393. <https://doi.org/10.3390/ani10081393>
35. Johnson, A. C., & Wynne, C. D. L. (2024). Comparison of the efficacy and welfare of different training methods in stopping chasing behavior in dogs. *Animals*, 14(18), Article 2632. <https://doi.org/10.3390/ani14182632>
36. Keeling, L. J., Winckler, C., Hintze, S., & Forkman, B. (2021). Towards a positive welfare protocol for cattle: A critical review of indicators and suggestion of how we might proceed. *Frontiers in Animal Science*, 2, Article 753080. <https://doi.org/10.3389/fanim.2021.753080>
37. Krueger, J. C., Boecker, M., Gauggel, S., Bleich, A., & Tolba, R. H. (2023). Using photographs for rating severity degrees of clinical appearance in research mice enables valid discrimination of extreme but not mild and moderate conditions: A pilot study. *PLOS ONE*, 18(11), Article e0287965. <https://doi.org/10.1371/journal.pone.0287965>

38. Laber, K., Newcomer, C. E., Decelle, T., Everitt, J. I., Guillen, J., & Brønstad, A. (2016). Recommendations for addressing harm-benefit analysis and implementation in ethical evaluation: Report from the AALAS-FELASA Working Group on Harm-Benefit Analysis - Part 2. *Laboratory Animals*, 50(1, Suppl.), 21-42. <https://doi.org/10.1177/0023677216642397>
39. Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2), 443-478. <https://doi.org/10.1162/003355397555253>
40. LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155-184. <https://doi.org/10.1146/annurev.neuro.23.1.155>
41. Leonardi, R. J., Vick, S.-J., & Dufour, V. (2012). Waiting for more: The performance of domestic dogs (*Canis familiaris*) on exchange tasks. *Animal Cognition*, 15(1), 107-120. <https://doi.org/10.1007/s10071-011-0437-y>
42. Li, Y., Zeng, J., Zhang, J., Yue, C., Zhong, W., Liu, Z., Feng, Q., & Luo, M. (2018). Hypothalamic circuits for predation and evasion. *Neuron*, 97(4), 911-924.e5. <https://doi.org/10.1016/j.neuron.2018.01.005>
43. Lindsay, S. R. (2005). *Handbook of applied dog behavior and training, volume three: Procedures and protocols*. Blackwell Publishing.
44. Marschark, E. D., & Baenninger, R. (2002). Modification of instinctive herding dog behavior using reinforcement and punishment. *Anthrozoös*, 15(1), 51-68. <https://doi.org/10.2752/089279302786992685>
45. Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55-73). Erlbaum.
46. McLean, A. N., & McGreevy, P. D. (2010). Ethical equitation: Capping the price horses pay for human glory. *Journal of Veterinary Behavior*, 5(4), 203-209. <https://doi.org/10.1016/j.jveb.2010.04.003>
47. Mellor, D. J. (2016). Updating animal welfare thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living." *Animals*, 6(3), Article 21. <https://doi.org/10.3390/ani6030021>
48. Mellor, D. J., Beausoleil, N. J., Littlewood, K. E., McLean, A. N., McGreevy, P. D., Jones, B., & Wilkins, C. (2020). The 2020 Five Domains Model: Including human-animal interactions in assessments of animal welfare. *Animals*, 10(10), Article 1870. <https://doi.org/10.3390/ani10101870>

49. Mendl, M., Burman, O. H. P., Parker, R. M. A., & Paul, E. S. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, 118(3-4), 161-181. <https://doi.org/10.1016/j.applanim.2009.02.023>
50. Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, 277(1696), 2895-2904. <https://doi.org/10.1098/rspb.2010.0303>
51. Moberg, G. P. (2000). Biological response to stress: Implications for animal welfare. In G. P. Moberg & J. A. Mench (Eds.), *The biology of animal stress: Basic principles and implications for animal welfare* (pp. 1-21). CABI Publishing.
52. Moseby, K. E., Cameron, A., & Crisp, H. A. (2012). Can predator avoidance training improve reintroduction outcomes for the greater bilby in arid Australia? *Animal Behaviour*, 83(4), 1011-1021. <https://doi.org/10.1016/j.anbehav.2012.01.023>
53. Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews Neuroscience*, 8(7), 536-546. <https://doi.org/10.1038/nrn2174>
54. Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., & Mendl, M. T. (2020). Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neuroscience and Biobehavioral Reviews*, 108, 749-770. <https://doi.org/10.1016/j.neubiorev.2019.11.014>
55. Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, 33(1), 56-72. <https://doi.org/10.1038/sj.npp.1301555>
56. Rodríguez, W., Bouzas, A., & Orduña, V. (2018). Temporal discounting of aversive consequences in rats. *Learning and Behavior*, 46(1), 38-48. <https://doi.org/10.3758/s13420-017-0279-9>
57. Rossier, D., La Franca, V., Salemi, T., Natale, S., & Gross, C. T. (2021). A neural circuit for competing approach and defense underlying prey capture. *Proceedings of the National Academy of Sciences*, 118(15), Article e2013411118. <https://doi.org/10.1073/pnas.2013411118>
58. Salgirli, Y., Schalke, E., Boehm, I., & Hackbarth, H. (2012). Comparison of learning effects and stress between 3 different training methods (electronic training collar, pinch collar and quitting signal) in Belgian Malinois police dogs. *Revue de Medecine Veterinaire*, 163(11), 530-535.

59. Sargisson, R. J., & McLean, I. G. (2021). Commentary: Efficacy of dog training with and without remote electronic collars vs. a focus on positive reinforcement. *Frontiers in Veterinary Science*, 8, Article 629746. <https://doi.org/10.3389/fvets.2021.629746>
60. Schuck-Paim, C., & Alonso, W. J. (2025). Welfare Footprint Framework: Methodological foundations and quantitative assessment guidelines. OSF. <https://doi.org/10.17605/osf.io/94bxs>
61. Shang, C., Liu, A., Li, D., Xie, Z., Chen, Z., Huang, M., Li, Y., Wang, Y., Shen, W. L., & Cao, P. (2019). A subcortical excitatory circuit for sensory-triggered predatory hunting in mice. *Nature Neuroscience*, 22(6), 909-920. <https://doi.org/10.1038/s41593-019-0405-4>
62. Shier, D. M., & Owings, D. H. (2006). Effects of predator training on behavior and post-release survival of captive prairie dogs (*Cynomys ludovicianus*). *Biological Conservation*, 132(1), 126-135. <https://doi.org/10.1016/j.biocon.2006.03.020>
63. Tetzlaff, S. J., Sperry, J. H., & DeGregorio, B. A. (2019). Effects of antipredator training, environmental enrichment, and soft release on wildlife translocations: A review and meta-analysis. *Biological Conservation*, 236, 324-331. <https://doi.org/10.1016/j.biocon.2019.05.054>
64. Vanderveldt, A., Oliveira, L., & Green, L. (2016). Delay discounting: Pigeon, rat, human - Does it matter? *Journal of Experimental Psychology: Animal Learning and Cognition*, 42(2), 141-162. <https://doi.org/10.1037/xan0000097>
65. Vieira de Castro, A. C., Fuchs, D., Morello, G. M., Pastur, S., de Sousa, L., & Olsson, I. A. S. (2020). Does training method matter? Evidence for the negative impact of aversive-based methods on companion dog welfare. *PLOS ONE*, 15(12), Article e0225023. <https://doi.org/10.1371/journal.pone.0225023>
66. Wang, W., Zhong, Y., Tan, R., Wang, M., Liu, J., Wang, D., Wang, H., Li, Y., Li, G., Yang, J., Wang, P., Wu, J., Zhang, J.-X., Wang, C.-Z., Shi, H., Ma, M., Yu, Y., & Zhang, Y.-F. (2025). A midbrain-to-ventral-striatum dopaminergic pathway orchestrates odor-guided insect predation in mice. *Proceedings of the National Academy of Sciences*, 122(41), Article e2514847122. <https://doi.org/10.1073/pnas.2514847122>
67. Weary, D. M., & Robbins, J. A. (2019). Understanding the multiple conceptions of animal welfare. *Animal Welfare*, 28(1), 33-40. <https://doi.org/10.7120/09627286.28.1.033>
68. West, R., Letnic, M., Blumstein, D. T., & Moseby, K. E. (2018). Predator exposure improves anti-predator responses in a threatened

- mammal. *Journal of Applied Ecology*, 55(1), 147-156.
<https://doi.org/10.1111/1365-2664.12947>
69. Wolfensohn, S., Sharpe, S., Hall, I., Lawrence, S., Kitchen, S., & Dennis, M. (2015). Refinement of welfare through development of a quantitative system for assessment of lifetime experience. *Animal Welfare*, 24(2), 139-149. <https://doi.org/10.7120/09627286.24.2.139>
70. Würbel, H. (2017). More than 3Rs: The importance of scientific validity for harm-benefit analysis of animal research. *Lab Animal*, 46(4), 164-166. <https://doi.org/10.1038/lab.1220>
71. Yeates, J. W. (2011). Is "a life worth living" a concept worth having? *Animal Welfare*, 20(3), 397-406. <https://doi.org/10.1017/S0962728600002955>
72. Yeates, J. W. (2012). Quality time: Temporal and other aspects of ethical principles based on a "life worth living." *Journal of Agricultural and Environmental Ethics*, 25(4), 607-624. <https://doi.org/10.1007/s10806-011-9317-3>
73. Yi, R., Gatchalian, K. M., & Bickel, W. K. (2006). Discounting of past outcomes. *Experimental and Clinical Psychopharmacology*, 14(3), 311-317. <https://doi.org/10.1037/1064-1297.14.3.311>
74. Yu, H., Xiang, X., Chen, Z., Wang, X., Dai, J., Wang, X., Huang, P., Zhao, Z.-D., Shen, W. L., & Li, H. (2021). Periaqueductal gray neurons encode the sequential motor program in hunting behavior of mice. *Nature Communications*, 12, Article 6523. <https://doi.org/10.1038/s41467-021-26852-1>
75. Zhao, Z.-D., Chen, Z., Xiang, X., Hu, M., Xie, H., Jia, X., Cai, F., Cui, Y., Chen, Z., Qian, L., Liu, J., Shang, C., Yang, Y., Ni, X., Sun, W., Hu, J., Cao, P., Li, H., & Shen, W. L. (2019). Zona incerta GABAergic neurons integrate prey-related sensory signals and induce an appetitive drive to promote hunting. *Nature Neuroscience*, 22(6), 921-932. <https://doi.org/10.1038/s41593-019-0404-5>
76. Zhao, Z.-D., Zhang, L., Xiang, X., Kim, D., Li, H., Cao, P., & Shen, W. L. (2023). Neurocircuitry of predatory hunting. *Neuroscience Bulletin*, 39(5), 817-831. <https://doi.org/10.1007/s12264-022-01018-1>
77. Ziv, G. (2017). The effects of using aversive training methods in dogs: A review. *Journal of Veterinary Behavior*, 19, 50-60. <https://doi.org/10.1016/j.jveb.2017.02.004>