# The North Setpoint

Why Grievance Survives the Reduction of Suffering

Flyxion
*Independent Researcher*

August 2026

**Abstract**

This essay develops a theory of grievance under conditions of sustained improvement. Standard hedonic-adaptation accounts explain how changes in circumstance can be followed by a return toward an affective baseline. The argument developed here concerns a different phenomenon: as conditions improve, the reference frame against which discrepancies are evaluated may itself migrate. Severe harms can become rarer while increasingly small departures from an adapted norm acquire greater relative salience. I call this process *grievance-scale renormalization* and use the term *North Setpoint* for the moving evaluative reference around which it occurs.

The resulting model distinguishes absolute hedonic condition from relative discrepancy, grievance from suffering, and the formation of a grievance from its eventual expression. It also separates several mechanisms that can otherwise appear identical from the outside: migration of the reference point, contraction of the experienced range, reallocation of attention as catastrophic problems disappear, and changes in the social threshold for expressing dissatisfaction. The reference point itself is further decomposed into components drawn from personal history, social comparison, and imagined counterfactual possibility, and a distinction is introduced between movement that is epistemically and normatively warranted and movement that merely describes what a distorted or captured comparison process happens to produce. Consequently, the frequency or intensity of observable complaint cannot be treated as a monotonic proxy for the amount of absolute suffering in a person or population.

David Pearce's proposal for information-sensitive gradients of bliss provides the limiting case. If negative valence were abolished while fine-grained discrimination among better and worse states remained, would grievance disappear with suffering, or would evaluative discrepancy simply migrate northward into an entirely positive hedonic range? The essay argues for the latter possibility while rejecting the reactionary inference that increasingly fine grievances are therefore trivial, decadent, or illegitimate. A civilization may become increasingly sensitive to correctable discrepancies precisely because catastrophic suffering has receded. The deeper design problem is therefore not necessarily to abolish discrimination, dissatisfaction, or information about what could be better, but to separate their informational function from the agony through which biological systems have historically made such information difficult to ignore. The essay closes by showing that a grievance can, in principle, retain this informational function even as its phenomenological cost is driven toward zero, giving the recurring formulation *keep the information, lose the agony* a precise, operational sense rather than only a memorable one.

# 1 A Mountain Made of a Molehill

Consider a public dispute over a single word. Someone uses a term another person considers inappropriate, a correction follows, and the correction is interpreted as accusation rather than clarification. The response to the correction then becomes evidence of character in its own right: friends choose sides, screenshots circulate, and a disagreement that began over a handful of characters of language becomes, within a few exchanges, a referendum on whether one participant is cruel, dishonest, untrustworthy, dangerous, or morally compromised in some more general and lasting sense.

Nothing in this example requires deciding who was right. The disputed word may genuinely matter, the correction may have been necessary, and the response may have revealed something real about the people involved. The point is not that linguistic or interpersonal grievances are imaginary simply because they concern objects smaller than a burned house or a broken bone, and it would be a mistake to read what follows as an argument that they are. The point is the extraordinary amount of evaluative resolution that can become concentrated on a small discrepancy once sufficiently larger threats have receded from ordinary life, so that a dispute occupying a few square inches of semantic territory can absorb the entire evaluative weight that a starving or endangered community would have had to reserve for matters of survival.

For much of human history, a lasting social rupture might accompany a stolen herd, a destroyed harvest, the loss of shelter, physical assault, exile, or a direct threat to one's family's survival, and grievances of a smaller magnitude simply had nowhere comparable to register: there was too much competing for attention at the top of the scale. In affluent and comparatively secure environments, by contrast, relationships can fracture over events whose absolute material consequences would scarcely have registered against those older backgrounds, and this is not obviously evidence of anything having gone wrong. It may equally be evidence that something has gone right, in the sense that the scale itself has been permitted to recalibrate around a narrower and gentler range of experience.

The tempting interpretation of this pattern is moral decline: people have become spoiled, and having escaped serious problems, they manufacture trivial ones to fill the space that hardship used to occupy. The opposite interpretation is equally tempting and no more obviously correct: contemporary grievances are simply more refined perceptions of harms that earlier societies lacked the security, vocabulary, or institutional capacity to notice, so that what looks like triviality is really precision. Neither interpretation is required by anything argued here, and part of the purpose of this essay is to show that the underlying phenomenon can be described without committing to either one.

The narrower puzzle, stated independently of any verdict on decline or refinement, is this:

Why does increasing safety not produce a proportionate disappearance of grievance?

If severe suffering declines by an order of magnitude, why does ordinary experience not necessarily contain an order of magnitude less dissatisfaction along with it? Why can a population become safer, healthier, wealthier, and more protected while continuing, apparently without contradiction, to produce a dense field of complaints, conflicts, and perceived discrepancies, sometimes concerning matters that would have seemed too small to register a generation earlier?

The answer developed here is that suffering and grievance do not occupy a single scale, however natural it is to speak as though they did. A grievance is not merely an absolute quantity of pain reported outward into the world; it is also, and perhaps primarily, an evaluation of a condition against a reference frame that the evaluator carries with them and updates continuously. If that

reference frame itself changes as conditions improve, then the disappearance of large negative events need not entail the disappearance of negative comparison, because comparison depends on the relation between an experience and a standard, and both terms of that relation are free to move.

The mountain can shrink, in other words, while the measuring instrument used to detect mountains becomes finer at the same time, and it is this second, usually invisible adjustment, the recalibration of the instrument rather than the shrinking of the object, that the remainder of this essay tries to make explicit.

## 2 The Ordinary Treadmill, and Why This Isn't That

The familiar hedonic-treadmill argument concerns the adaptation of affect. Circumstances improve or deteriorate, subjective well-being moves in response, and over time some portion of that movement dissipates as the organism adapts back toward a characteristic range, so that a lottery win or a debilitating injury each produces less lasting change in reported happiness than naive intuition would predict. In schematic form,

$$\text{circumstances change} \longrightarrow H_t \text{ changes} \longrightarrow H_t \text{ tends toward a characteristic range.}$$

The strongest versions of setpoint theory treat this characteristic range as comparatively stable across the life course, fixed largely by temperament and biology. More qualified versions allow the range itself to move with persistent changes in health, environment, chronic stress, illness, relationships, or other durable conditions, and it is this more qualified picture that the present argument takes as its starting point rather than its target.

The present argument concerns a different variable than either version of the treadmill addresses directly. Suppose the hedonic baseline itself actually improves, so that instead of repeatedly returning to the same fixed point after each disturbance, the organism or population moves north in a lasting way:

$$N_{t+1} > N_t.$$

The question this essay asks is then not whether improvement survives adaptation. Assume that it does, and grant the most generous possible reading of the evidence: assume that tomorrow's population really is happier than today's, and that today's really is happier than yesterday's, with no illusion or measurement artifact involved. The question that remains, even after that generous assumption is granted in full, is what happens to the *reference frame* used to distinguish better from worse inside the newly improved range.

A treadmill account says, in effect, that

$$\text{the environment improves but affect returns.}$$

The North Setpoint asks a different question, compatible with the treadmill account but not reducible to it:

$$\text{affect improves and the evaluative scale moves with it.}$$

These are different levels of explanation, addressed to different failure modes of naive optimism. The first concerns the persistence of hedonic improvement itself, and its natural adversary is the worry that gains will simply evaporate. The second concerns the persistence of discrimination after

improvement has genuinely occurred and persisted, and its natural adversary is the different worry that even durable, unambiguous improvement might fail to produce a corresponding decline in the felt sense that something remains wrong.

Nothing in the North Setpoint requires happiness gains to be illusory, exaggerated, or self-cancelling. Indeed, its most interesting and most difficult case begins by granting that they are entirely real, stipulating the strongest possible version of genuine improvement, and asking what an evaluative system built to detect discrepancy would still be doing once that improvement had fully arrived.

## 3 Grievance-Scale Renormalization

Let $H_t$ denote experienced hedonic condition at time $t$. Define a locally adapted reference point over some relevant temporal window $\tau$ as

$$N_t = \mathbb{E}[H]_{t-\tau:t}\,.$$

This expression is deliberately simple, and it is worth being explicit about how much it leaves out. Actual reference formation may be weighted toward recent or salient events rather than uniformly averaged, nonlinear in ways that give disproportionate influence to extremes, socially mediated through comparison with visible peers rather than purely autobiographical, dependent on long-term memory in ways that a short trailing window cannot capture, domain-specific so that income, health, and relationships each carry their own separately adapting reference points, and asymmetric between gains and losses in the manner familiar from prospect theory. None of these refinements is incorporated here, because for present purposes the essential feature is only that recent experience contributes to the standard against which present experience is evaluated, and the simplest possible functional form suffices to make that point precisely.

Under sustained improvement,

$$N_{t+1} > N_t.$$

Now let grievance depend not solely on the absolute value of an event $H_i$, but also on its discrepancy from the adapted reference point:

$$G_i(t) = g(N_t - H_i,\, H_i,\, C_i,\, M_t, \ldots)\,,$$

where $C_i$ may represent contextual variables such as perceived fairness, agency, or social meaning attached to the event, and $M_t$ accumulated memory or comparison information that colors how the discrepancy is interpreted once it is registered. The first argument, $N_t - H_i$, is reference-relative discrepancy, the quantity by which an event falls short of what the system has come to expect. The second, $H_i$, retains absolute hedonic information, so that the function $g$ is not forced to discard the very quantity whose fate is at issue.

Keeping both terms in the function is important, and it is the feature that distinguishes this model from a purely relativistic account of well-being. The theory does not require that relative comparison erase absolute magnitude, only that it be capable of contributing independently to the resulting grievance. A model that collapsed everything onto the single term $N_t - H_i$ would predict that no amount of absolute improvement could ever reduce grievance, which is too strong a claim and one this essay does not make; a model that used only $H_i$ would predict that grievance tracks absolute suffering exactly, which is the very claim under dispute. The two-term structure is what allows both possibilities, less absolute suffering with unchanged relative grievance, and less

absolute suffering with genuinely less relative grievance, to remain live outcomes depending on how the reference point itself moves.

Suppose the relevant experienced range extends from a local floor $H_{\min}(t)$ to the reference point $N_t$. A simple normalized measure of relative severity is

$$\rho_i(t) = \frac{N_t - H_i}{N_t - H_{\min}(t)}.$$

If the lower tail of experience moves northward, so that the very worst states a population now encounters are themselves better than the worst states it used to encounter, then

$$H_{\min}(t+1) > H_{\min}(t).$$

The denominator $N_t - H_{\min}(t)$, the total span between the reference point and the floor, can contract as a result. An event of comparatively small absolute magnitude can thereby occupy a large fraction of the remaining negative range, simply because the range itself has narrowed around it, not because the event has changed in any absolute sense.

This is grievance-scale renormalization. What once occupied the bottom one percent of an enormous experiential range may, after the disappearance of more severe states from that range, come to occupy the bottom twenty percent of a much narrower one. The event did not become absolutely worse by any measure external to the ranking; only its position within the active comparison space changed, because the space around it contracted.

Hence the central paradox of this essay, stated as compactly as the argument allows:

$$\boxed{\text{less absolute suffering } \not\Rightarrow \text{ proportionally less experienced grievance.}}$$

The implication must not be overstated, and it is worth pausing on the ways in which it could be overstated before moving on. The model does *not* predict that grievance must remain constant as suffering falls, and still less does it predict that grievance must increase as a matter of law. A population may experience both dramatically less suffering and dramatically fewer grievances at the same time, if the reference point and the floor happen to move together in a way that leaves relative severity roughly unchanged or even reduced. Nothing in the mathematics forces the pessimistic reading; it only removes the guarantee of the optimistic one.

The narrower claim, the one this essay is actually prepared to defend, is more defensible precisely because it is narrower:

$$\boxed{\text{grievance frequency is not a monotonic proxy for absolute suffering}}$$

once the reference frame against which grievances are generated is itself adaptive, as the evidence from adaptation research suggests it generally is. A society reporting many complaints is therefore not necessarily suffering more than a society reporting few, and a decline in complaints does not, by itself, establish a decline in suffering either. The two variables can move together, as intuition usually assumes they must, but they need not, and the gap between what intuition assumes and what the model permits is exactly the space this essay is trying to occupy.

## 4 Comparison Populations

The reference point introduced above was deliberately simple. Writing $N_t = \mathbb{E}[H]_{t-\tau:t}$ made it possible to isolate the central mechanism of grievance-scale renormalization without immediately

multiplying its causes, so that sustained improvement could be shown, on its own, to change the range against which present conditions are evaluated. But a person's evaluative north is almost certainly not constructed from autobiographical experience alone, and treating it as though it were leaves out most of what actually populates an ordinary reference frame. Human beings compare their circumstances not merely with what they themselves experienced yesterday, but with what other people appear to experience, with what they have been told is normal, with what they believe ought to be possible, and with counterfactual states that may never yet have been instantiated at all. The operative reference point is therefore better understood as a composite object assembled from several comparison populations rather than as a moving average over one experienced trajectory.

It is worth stating this generalization explicitly rather than leaving it as an unannounced replacement of the earlier model, since the two are not competitors but stages of the same construction. Define a generalized reference operator

$$N_i(t) = \mathcal{N}(\mathcal{H}_i(t),\, \mathcal{S}_i(t),\, \mathcal{C}_i(t))\,,$$

where $\mathcal{H}_i(t)$ is agent $i$'s experienced history up to time $t$, $\mathcal{S}_i(t)$ the states observed in some surrounding social population, and $\mathcal{C}_i(t)$ the set of counterfactual states the agent currently regards as reachable. The moving-average reference point of the previous section is then not a separate mechanism sitting alongside this one, but the special case obtained by suppressing the social and counterfactual arguments entirely and letting $\mathcal{N}$ act as a simple trailing expectation over $\mathcal{H}_i(t)$ alone,

$$N_t = \mathbb{E}[H]_{t-\tau:t} \quad \Longleftarrow \quad \mathcal{N}(\mathcal{H}_i(t), \varnothing, \varnothing)\,.$$

Every refinement introduced from this point forward, comparison populations, attainability, justificatory weight, and the sections still to come, is best understood as a further specification of what $\mathcal{N}$, $\mathcal{S}_i$, or $\mathcal{C}_i$ are actually built from, rather than as an independent addition standing beside the original renormalization argument. The mathematics of the essay is cumulative in this sense: nothing introduced later discards what came before, and everything introduced later can be read as an instantiation of the same underlying operator with a richer set of arguments supplied to it.

A minimal instantiation of this operator separates three such sources. Let $N_i^{\text{self}}(t)$ denote the reference distribution generated from agent $i$'s own recent experience, let $N_i^{\text{social}}(t)$ denote the reference supplied by observations of other agents, and let $N_i^{\text{cf}}(t)$ denote a counterfactual reference formed from states the agent believes could reasonably obtain. Then the effective north setpoint may be written

$$N_i(t) = \alpha_i N_i^{\text{self}}(t) + \beta_i N_i^{\text{social}}(t) + \gamma_i N_i^{\text{cf}}(t), \qquad \alpha_i + \beta_i + \gamma_i = 1,$$

with the coefficients understood not as fixed psychological constants but as context-sensitive weights that can themselves shift as an agent's information environment changes. A person whose information environment is narrow may have $\alpha_i$ close to one, because recent personal experience supplies most of the available comparison class and there is little else to weigh it against. Another person inhabiting a dense social information environment may carry a much larger $\beta_i$, constantly recalibrating against visible peers, while an imaginative, technically informed, or politically aspirational agent may place considerable weight on $\gamma_i$, attending not merely to how things are elsewhere but to how much better they could already be, given what is now known to be possible.

This changes the interpretation of grievance in an important way, one the single-term model of the previous section could not capture on its own. It is no longer necessary for an agent's absolute condition $H_i(t)$ to deteriorate, or even for that agent's recent experiential history to improve, in order for the discrepancy

$$\Delta_i(t) = N_i(t) - H_i(t)$$

to increase. Holding $H_i(t)$ fixed and ignoring, for the moment, any change in the weights themselves gives

$$\dot{\Delta}_i(t) = \alpha_i \dot{N}_i^{\text{self}} + \beta_i \dot{N}_i^{\text{social}} + \gamma_i \dot{N}_i^{\text{cf}}.$$

A person can therefore become more dissatisfied with an unchanged condition because the social comparison distribution has moved north around them, because previously invisible better conditions have become visible through wider information access, or because some newly discovered possibility has transformed what counts as an attainable alternative. The change need not be envy in any pejorative sense, and it need not involve any deterioration of mood or any worsening of the underlying condition being evaluated. It is a change in the coordinate system within which the same condition is being measured, arriving from outside the agent's own experiential history entirely.

The counterfactual term is especially important because it prevents the theory from collapsing into a simple account of social comparison, which would be a much thinner and less interesting claim than the one this essay is trying to make. Some grievances have no existing comparison population at all. A person may object to a technical limitation, an institutional arrangement, an architectural inconvenience, or a social practice not because someone else already enjoys a demonstrably better alternative, but because the alternative has become conceivable in the abstract, perhaps for the first time. Once a previously invisible possibility enters the reachable state space, the present condition acquires a new neighbor against which it can be evaluated, even where no one anywhere has yet occupied that neighboring state. In this sense invention can manufacture grievances without manufacturing any prior harm to license them: the first person to understand that some burden is unnecessary may become dissatisfied before anyone has actually escaped it, sometimes long before.

This suggests that the North Setpoint should ultimately be treated less as a single scalar thermostat than as an adaptive reference distribution over reachable and imaginable states, continuously reassembled from self, society, and possibility. The better a system becomes at representing alternatives, whether through richer social information, more capable institutions, or more powerful imagination, the larger this distribution may become, and therefore the more ways its present condition can be found locally suboptimal even while remaining, by any external measure, excellent. That observation becomes particularly consequential once the worst conditions themselves begin to disappear, because removing large discrepancies does not merely improve the distribution in the way ordinary intuition expects. It also changes which smaller discrepancies are available, for the first time, to occupy the attention that catastrophe used to monopolize.

## 5 The Vanishing Catastrophe and the Expanding Error Budget

Grievance-scale renormalization, as developed above, explains why the same discrepancy can acquire greater relative salience as the surrounding range narrows, and the comparison-population extension explains why the reference point itself can move even when an agent's own experience has not changed at all. Both mechanisms, however, remain accounts of evaluation: they concern what counts as bad, not what gets noticed and acted upon. A third, complementary mechanism operates at the level of attention and repair capacity rather than at the level of evaluation, and it does not depend on any psychological adaptation whatsoever.

Large harms do not simply sit at the bottom of the hedonic scale like inert data points waiting to be measured against a reference. They actively consume the evaluative and practical bandwidth of whatever system is dealing with them. A famine, a war, or an epidemic occupies not only the sufferer's own attention but the attention of institutions, caregivers, journalists, and observers whose collective capacity to notice and respond to problems is finite at any given moment. Repair capacity, the resources an individual or a society devotes to fixing what is wrong, gets allocated disproportionately toward the largest and most urgent discrepancies in view, and smaller ones simply never rise high enough in the queue to receive attention, however real they may already be.

This can be made explicit with a simple allocation model. Let $B_t$ denote the total evaluative, attentional, and repair budget available to a system, whether an individual mind or a civilization's collective institutions, at time $t$. Assume, plausibly, that $B_t$ is bounded: no system can attend to every live discrepancy simultaneously, however small. Suppose further that this budget is spent on discrepancies roughly in descending order of the relative severity measure $\rho_i(t)$ introduced earlier, with the largest, most urgent problems drawing resources first and the remainder cascading down the ranking until the budget is exhausted. Under this allocation rule, a discrepancy far down the ranking may never receive meaningful attention at all, not because it is judged unimportant on reflection, but because it never survives long enough in the queue to be judged at all.

When a large discrepancy vanishes, whether through the end of a war, the eradication of a disease, or the closing of some other catastrophic gap, the budget that discrepancy had been consuming is not simply idle afterward. It is redeployed to whatever remains below it in the ranking. This produces a structural, rather than a purely psychological, explanation for why smaller discrepancies become newly salient once larger ones disappear: the attentional and institutional resources needed to notice and address them were, until that point, monopolized by problems of greater urgency, and their release is itself sufficient to surface what had been sitting, entirely unattended, further down the list.

This mechanism is independent of, and compounds with, grievance-scale renormalization rather than duplicating it. Renormalization changes what counts as bad by moving the reference point and the floor; the error-budget mechanism changes what gets noticed and acted upon at all, holding the reference point fixed. Even a population whose evaluative standards had somehow stopped adapting entirely would still, on this account, discover a wave of newly salient smaller problems the moment its larger ones were resolved, purely because the capacity to attend to them had finally been freed. The two mechanisms working together, one shifting the scale and the other redirecting the spotlight along it, make the appearance of newly prominent small grievances following large-scale improvement considerably more overdetermined than either mechanism alone would suggest, and considerably harder to read as evidence of decline.

The metaphor is borrowed loosely from reliability engineering, where an error budget denotes the total tolerable deviation from perfect performance that a system is permitted before it must divert resources from new development to repair. Here the analogous budget is not banked once the largest failure modes are eliminated; it is spent finding the next ones. A civilization that has exhausted its catastrophes has not exhausted its capacity for discovering what is wrong with itself. It has, for the first time, freed the capacity needed to look.

## 6 Three Ways a Small Problem Becomes Large

The preceding sections have accumulated three distinct mechanisms by which a discrepancy of fixed absolute magnitude can come to occupy an outsized share of a system's evaluative attention. They have arrived sequentially, each introduced to solve a specific gap left by the one before, and it

is worth pausing before proceeding further to state plainly that they are three mechanisms rather than three descriptions of one.

The first is renormalization of the active range. As the floor $H_{\min}(t)$ rises and the span between floor and reference contracts, an event of unchanged absolute severity occupies a larger fraction of what remains, captured by

$$\rho_i(t) = \frac{N_t - H_i}{N_t - H_{\min}(t)}.$$

Nothing about this mechanism requires the reference point $N_t$ itself to move for any reason beyond the ordinary adaptation already described; it is a claim about the shrinking denominator, not about the numerator.

The second is migration of the reference point itself, developed through the generalized operator $\mathcal{N}(\mathcal{H}_i, \mathcal{S}_i, \mathcal{C}_i)$ and its self/social/counterfactual instantiation. Here the discrepancy

$$\Delta_i(t) = N_i(t) - H_i(t)$$

can grow even while $H_i(t)$ and $H_{\min}(t)$ are both held fixed, simply because the social or counterfactual inputs to $N_i(t)$ have shifted. This mechanism operates on the numerator of $\rho_i(t)$ rather than its denominator, and it can operate entirely independently of any change in the experiential range itself.

The third is reallocation of attention. Even holding both $\rho_i(t)$ and $\Delta_i(t)$ fixed, a discrepancy's attentional rank $A_i^{\text{attn}}(t)$, its position in the queue for the finite budget $B_t$, can rise simply because larger discrepancies that had been consuming that budget have disappeared. This mechanism concerns neither the scale used to measure severity nor the target against which severity is measured, but the separate question of whether the system doing the measuring ever gets around to attending to a given discrepancy at all.

These three quantities can be gathered into a single salience function without collapsing the distinction between them,

$$\text{Salience}_i(t) = \mathcal{S}\big(\rho_i(t),\, \Delta_i(t),\, A_i^{\text{attn}}(t)\big)\,,$$

and the discipline the essay has tried to maintain throughout is precisely that this function takes three arguments rather than one. A discrepancy can become salient because $\rho_i(t)$ has risen while $\Delta_i(t)$ and $A_i^{\text{attn}}(t)$ are unchanged; because $\Delta_i(t)$ has risen while $\rho_i(t)$ and $A_i^{\text{attn}}(t)$ are unchanged; or because $A_i^{\text{attn}}(t)$ has risen while the other two are unchanged. In practice the three usually move together, since the same episode of large-scale improvement tends to raise the floor, shift the reference population, and free attentional budget all at once, which is exactly why ordinary experience tends to blur them into a single undifferentiated sense that small things have started to matter more. But the essay's claim has never depended on their moving together, only on the observation that none of the three logically requires either of the other two:

$$\boxed{\text{renormalization} \neq \text{reference migration} \neq \text{attention reallocation.}}$$

Keeping the three apart matters for more than definitional tidiness. Renormalization is corrected, if it needs correcting at all, by widening the frame within which severity is judged. Reference migration is corrected by auditing the comparison populations feeding $N_i(t)$, a question taken up directly in a later section. Attention reallocation is corrected, if it is corrected, simply by continuing to spend the freed budget, since the mechanism describes a redistribution of finite capacity rather than any distortion in judgment. A single observed increase in complaint about small things could,

in a given case, be traceable to any one of the three, to any two, or to all three at once, and mistaking which is operating is precisely the kind of error the rest of this essay has tried to make it harder to commit.

## 7 A Second Illustration, Away from Suffering Entirely

The same structure appears in domains where suffering is not obviously at issue at all, which is useful because it shows that the mechanism under discussion is not a peculiarity of hedonic psychology but a more general property of bounded evaluative systems.

Imagine drinking inexpensive fruit juice every evening for years, and then one night having it replaced by competent table wine. At first the difference is enormous: the wine is aromatic, structured, dry, complex, and startlingly unlike the juice that had, until that moment, established the entire reference frame for an evening drink. Within a few months, however, wine becomes ordinary. Within a year or two, good wine becomes ordinary as well, and the relevant distinctions have migrated to acidity, oxidation, balance, serving temperature, storage conditions, vintage, region, or whether the olives served alongside it overwhelm something delicate in the glass, distinctions that would have been not merely unnoticed but literally meaningless to the earlier, juice-drinking version of the same person.

Someone at this later stage may eventually complain that an otherwise excellent wine was served a few degrees too warm. It would be strange to conclude from this complaint that the person therefore enjoys wine less than the person whose baseline remained fruit juice and who is delighted by anything fermented at all. The opposite is closer to the truth: the entire experiential range has moved upward, and what looks like increased fussiness is, from another angle, increased capacity for pleasure, now expressed as sensitivity to increasingly fine gradations within a region of experience that used to be entirely inaccessible.

What changed, in other words, was the resolution of discrimination, not merely its target. Let $Q_i$ denote absolute quality and $N_t^Q$ an adapted quality reference specific to this domain. Then dissatisfaction can arise from

$$N_t^Q - Q_i > 0$$

even where

$$Q_i \gg Q_{\text{old baseline}},$$

so that the very same bottle, judged against the person's own earlier standard, would have counted as a revelation, and judged against their current standard counts as a minor disappointment, without either judgment being mistaken on its own terms.

A rising floor does not eliminate discrimination; it relocates discrimination into increasingly fine structure, redirecting the same discriminative capacity that once distinguished juice from wine toward distinctions that only make sense once wine itself has become the baseline. The mechanism is therefore not intrinsically hedonic, and nothing about it depends on pain or pleasure specifically. Any bounded evaluative system capable of relative discrimination can exhibit something analogous, because the underlying requirement is only that the system rank states against a moving standard, and as coarse failures disappear from the available range, attention and judgment migrate toward whatever finer differences remain to be found.

This observation matters for a reason beyond the wine itself: it prevents the North Setpoint from becoming a disguised claim that humans secretly require suffering in order to function, a claim this

essay explicitly does not want to make and takes some care to block. No suffering of any kind is necessary for the wine to be judged slightly too warm. What is necessary is only that the system preserve an ordering over states within an already desirable region, and preserving that ordering is a cognitive achievement, not a hedonic deficit.

## 8 David Pearce and the Far North

David Pearce's abolitionist project provides an unusually strong limiting case for this argument, and it is worth engaging in some detail precisely because it is designed, by its own author, to be the hardest version of the problem rather than the easiest.[1] His proposal is not merely that humanity should become somewhat happier on average, or that medicine should continue its ordinary work of alleviating conventional psychiatric and physical suffering. The stronger aspiration, stated repeatedly and without qualification across his writing, is the technological abolition of involuntary suffering itself, achieved not by managing suffering more effectively but by removing the biological machinery that produces it.

A superficial objection immediately presents itself to any such proposal, and it is worth stating clearly in order to set it aside. If people adapt to improvement in the ordinary treadmill sense, perhaps they would simply become miserable again at a higher material level once the machinery of suffering had, in principle, been switched off; raise the floor and the hedonic treadmill will simply recreate the old problem one rung higher on the ladder, leaving the abolitionist project no better off than any other program of material improvement.

Pearce's more ambitious formulation makes that objection inadequate as a response to what he is actually proposing. In his discussion of *information-sensitive gradients of bliss* [2], the proposal is not simply to increase $H_{\text{min}}$ while leaving an otherwise unchanged pain and pleasure architecture in place underneath the new, higher floor. Instead, the information-bearing function that affect currently performs could itself be moved entirely into positive territory, so that the architecture generating discrepancy signals no longer has a negative branch at all. Differences in motivation, priority, error, success, urgency, and preference need not, on this proposal, be encoded by transitions between suffering and pleasure; they might instead be encoded entirely by gradients such as

$$\text{bliss}_1 < \text{bliss}_2 < \text{bliss}_3 < \cdots,$$

where every term in the sequence satisfies

$$\text{bliss}_i > 0 \qquad \forall i.$$

The lowest state available to such an organism could, by contemporary standards, be extraordinarily good, comparable perhaps to the peak experiences that presently occupy only the rarest moments of an ordinary human life, and yet still function, within that organism's own economy of motivation, as the state to be avoided in favor of something even better.

This is a much harder case for the North Setpoint than the wine example or the ordinary treadmill, and precisely because it is harder, it is also considerably more interesting. It removes the easy criticism that abolition merely changes what people suffer about, relabeling old pains as new pleasures without actually eliminating the underlying negative machinery. If the proposal works exactly as Pearce intends, the lower state in the sequence is not suffering wearing a disguise; it is genuinely, unambiguously positively valenced, by the same phenomenological criteria that would be used to identify positive valence anywhere else.

[1] Pearce, *The Hedonistic Imperative* [1].

The question becomes instead:

> What happens to discrepancy when the entire space in which discrepancy occurs has moved north?

If intelligent action still depends, as it plausibly must, on differences between states that the organism can register and act upon, then abolishing suffering cannot mean abolishing discrimination as such, on pain of abolishing the very capacity for goal-directed behavior that made the improvement worth pursuing in the first place. A system that cannot distinguish

$$H_A > H_B$$

cannot preferentially navigate toward $H_A$ on the basis of that distinction, and would, in an important functional sense, no longer be an agent with preferences at all, however pleasant its undifferentiated condition might be.

Thus the Far North does not abolish the problem of comparison; it purifies it, stripping away the agony that has historically accompanied comparison in biological organisms while leaving the comparative structure itself, and everything that depends on it, intact.

## 9 Four Things That Are Not the Same

Much of the confusion surrounding proposals like Pearce's follows from treating several distinct forms of negativity as though they were interchangeable, when careful attention shows that they are not, and that a proposal can eliminate one of them while leaving the others fully in place. At minimum, four distinctions are required to keep the discussion honest:

negative valence,

negative comparison,

negative judgment,

and

behavioral complaint.

Negative valence is phenomenological. It concerns what a state feels like from the inside, independent of any comparison to any other state, and it is the quantity that abolitionist projects like Pearce's target most directly and most literally for elimination.

Negative comparison is relational rather than phenomenological. It requires only that one state be ranked below another,

$$H_B < H_A,$$

and this ranking can obtain, as the numerical example below shows, even where both states are, in absolute phenomenological terms, wonderful.

Negative judgment introduces evaluation on top of mere ranking. It asserts something further, that

$$H_B \text{ is worse than it should be,}$$

or that

$$H_B \text{ ought to be replaced by } H_A,$$

a normative claim that a mere ranking does not by itself entail, since a system could register that $H_B < H_A$ without concluding that anything ought to be done about it.

Behavioral complaint, finally, is an expression generated downstream from evaluation of this kind, and it is the most contingent of the four. Even negative judgment does not guarantee complaint, because social cost, self-monitoring, opportunity, prevailing norms, personality, and a host of other constraints can intervene between an evaluation reached privately and any public or behavioral expression of it.

Pearce's abolitionist proposal requires the first category, negative valence, to disappear, or at minimum to be driven asymptotically toward disappearance. His own further requirement, that bliss remain information-sensitive rather than collapsing into a single undifferentiated maximum, requires something like the second category, negative comparison, to survive intact. If $A$ denotes action selection, then meaningful affective discrimination requires, schematically,

$$\frac{\partial A}{\partial H} \neq 0,$$

meaning that differences in hedonic condition must still be capable of making a difference to behavior, on pain of the organism losing all capacity for preference-guided action. But satisfying this requirement does not, by itself, determine whether negative judgment or behavioral complaint must also survive, and in what form; those two categories are left underdetermined by the bare requirement that comparison continue to function.

Consider an intentionally absurd numerical example, useful precisely because its absurdity makes the logical structure easier to see than a more realistic case would. Let contemporary peak human experience have the arbitrary value

$$H_{\text{peak,human}} = 100.$$

Now imagine two posthuman states,

$$H_A = 10{,}000$$

and

$$H_B = 9{,}999.$$

Both are, by the stipulated scale, fantastically better than the greatest pleasure presently available to any human being, by a factor of one hundred. Yet

$$H_B < H_A,$$

and the ranking is unambiguous even though both terms sit far above anything a human evaluator could presently experience. Suppose further that the organism's adapted reference point, formed by the same mechanism described in the previous section, has itself risen to

$$N_t \approx 10{,}000.$$

Then simultaneously

$$H_B > 0$$

and

$$H_B - N_t < 0.$$

There is no contradiction in holding both facts at once. The state can be extraordinarily good in absolute terms and suboptimal in relative terms at the very same moment, evaluated by the very same organism, without either evaluation being mistaken.

The cleaner question for the North Setpoint is therefore not whether suffering will mysteriously regenerate itself after suffering has been abolished, which would be a strange and largely empirical worry about the robustness of the abolitionist technology. It is instead the sharper conceptual question:

| Does eliminating suffering eliminate negative appraisal, or merely move its reference frame north? |
|---|

The distinction matters because a preference gradient is not automatically a pain gradient, and a negative comparison is not automatically negative phenomenology, even though ordinary language routinely treats the two as synonyms. Failure to preserve those distinctions makes the thought experiment trivial, since any negative-sounding word can then be used to declare the whole abolitionist project either self-refuting or already accomplished depending on the rhetorical needs of the moment. Preserving them makes the thought experiment difficult, in the more productive sense of forcing genuine argument rather than permitting a verbal shortcut in either direction.

## 10 Grievance Without Misery

The four-way distinction just established, between negative valence, negative comparison, negative judgment, and behavioral complaint, makes it possible to ask a sharper question than anything posed so far: what does grievance actually become as phenomenological suffering, the first of the four categories, approaches zero across an entire population, as Pearce's Far North supposes it eventually will? Ordinary experience offers little guidance here, because in ordinary experience grievance and misery are so tightly bound together that the two are rarely disentangled even in careful reflection, let alone in the vocabulary available for describing them.

Rather than define grievance implicitly through pain, as ordinary usage tends to, it is more useful to define it explicitly as an error signal concerning an attainable ordering, deliberately built from ingredients that do not presuppose suffering at any point in their construction. Let

$$G_i(t) = \Psi(N_t - H_i,\ A_i,\ J_i)\,,$$

where the first argument is the familiar reference-relative discrepancy from earlier sections, and the two new arguments carry the additional structure this section is after. Let $A_i$ denote attainability, the degree to which some superior state $H_j > H_i$ is genuinely reachable from $H_i$ through available action, rather than being merely a notional possibility with no path leading to it. Let $J_i$ denote justificatory weight, the degree to which the gap between $H_i$ and the reference point is attributed to a cause that could, or should, have been otherwise, rather than to brute and unavoidable structure that no agent or institution could reasonably be expected to have prevented.

This formalizes an intuition that ordinary language handles only clumsily: a grievance is not merely the observation that some state falls short of expectation, but the further claims that the shortfall is reachable and that someone or something bears some responsibility for it not yet having been closed. A state can register a large discrepancy $N_t - H_i$ and still generate almost no grievance if $A_i$ is low, because nothing can presently be done about it, or if $J_i$ is low, because the gap is attributed to impersonal and unavoidable causes rather than to a fixable failure. Conversely, a comparatively small discrepancy can generate intense grievance if both $A_i$ and $J_i$ are high, because the better state is visibly within reach and its absence looks, to the evaluator, like a choice rather than a fact of nature.

This distinction is not merely formal; it tracks a real historical pattern. Severe suffering from disease before the germ theory of infection was widely accepted often produced fatalism rather than grievance, not because the suffering itself was smaller in absolute terms, but because $A_i$ was effectively zero: there was, as far as anyone could tell, nothing to be done, and nothing to hold responsible. Once effective treatments existed, historically comparable suffering became newly and intensely grievance-laden, even in cases where $H_i$ itself had not changed at all, because $A_i$ had risen from near zero to substantial, and often $J_i$ alongside it, once a particular institution's failure to provide the now-available treatment became a live possibility to blame. Technological and institutional capacity, on this account, manufacture grievance directly, by raising attainability, entirely independent of any renormalization of the reference point $N_t$ itself.

The payoff of this formalization becomes visible once it is carried into the Far North. Suppose suffering, in the sense of negative valence, has genuinely been abolished, so that $H_i(t) > 0$ for every agent at every time, exactly as Pearce's proposal requires. Nothing in that stipulation forces $A_i$ or $J_i$ toward zero. An agent inhabiting an information-sensitive gradient of bliss can still identify neighboring states that are better than its own, can still, in principle, act to reach them, and can still, in a civilization sophisticated enough to design its own hedonic architecture in the first place, attribute the persistence of any remaining gap to a choice that some other agent or institution could in principle have made differently. Under those conditions, $G_i(t)$ can remain strictly positive even while $H_i(t)$ is enormous and, by any contemporary standard, ecstatic. This is grievance without misery: an evaluative, motivating dissatisfaction that carries no phenomenological sting at all, functioning purely as guidance rather than as agony.

This gives a cleaner and more constructive answer to the boxed question posed at the end of the preceding section. Eliminating suffering does not eliminate negative appraisal, not because negative appraisal is somehow indestructible, but because negative appraisal, properly formalized, never actually required negative valence as one of its inputs in the first place. It required discrepancy, attainability, and attribution, three quantities Pearce's proposal, on this reading, leaves fully intact even as it drives the fourth quantity, valence itself, asymptotically toward its floor of positivity.

## 11 From Punishment Signal to Guidance Signal

The reformulation of grievance just completed suggests a distinction the essay has been gesturing toward since its opening pages without quite naming directly: a grievance can become increasingly informational at the same time as it becomes decreasingly punitive, and these two properties, so tightly fused in biological experience that they are rarely even considered separable, are in fact independent dimensions of the same signal.

Separate the information a grievance carries from the phenomenological cost of carrying it. Let

$$I_G = I(G;\, X^* - X)$$

denote the information the grievance signal $G$ carries about some correctable discrepancy between an actual state $X$ and a superior attainable state $X^*$, in the ordinary information-theoretic sense of mutual information between the signal and the quantity it is tracking. Let

$$C_G = -\min(H_G, 0)$$

denote the negative phenomenological cost of that same signal, the magnitude of whatever suffering, if any, accompanies its occurrence, defined so that $C_G$ is zero whenever the associated hedonic state $H_G$ is non-negative and grows only as $H_G$ becomes negative.

Biological systems, shaped by selection pressures that favored signals difficult to ignore, typically bundle these two quantities together, so that

$$I_G > 0, \qquad C_G > 0$$

hold simultaneously as a matter of course: the same event that tells the organism something has gone wrong is, by design, also unpleasant to experience, and the unpleasantness is precisely what recruits attention and motivates correction. This bundling is so pervasive in ordinary experience that it is tempting to treat it as necessary rather than contingent, as though information about discrepancy could not be transmitted any other way.

But there is no logical requirement that useful information about a correctable discrepancy be accompanied by any phenomenological cost at all. Nothing in the definition of $I_G$ makes reference to $H_G$, and nothing in the definition of $C_G$ makes reference to whether the signal is informative. The two quantities are conceptually orthogonal even though biological evolution, for reasons of engineering expedience rather than logical necessity, has historically wired them together. The Far North, on this account, is best understood as the limiting case in which a civilization has learned to separate what evolution bundled, approaching the joint condition

$$I_G > 0, \qquad C_G \to 0.$$

This gives the essay's recurring refrain an actual optimization target rather than only a memorable formulation. *Keep the information, lose the agony* has, up to this point, functioned as a philosophically effective slogan, particularly in the discussion of the inversion that follows. Stated as $I_G > 0,\ C_G \to 0$, it becomes a criterion against which any proposed engineering of affect, Pearce's or otherwise, can actually be evaluated: a proposal succeeds to the degree that it drives $C_G$ toward its floor while holding $I_G$ away from zero, and it fails, whatever its other merits, if it purchases a lower $C_G$ at the cost of a collapsing $I_G$, producing not a Far North but a merely numb one, undifferentiated and no longer capable of guiding behavior toward anything in particular.

## 12 The Inversion

Darwinian affect uses suffering partly because suffering is difficult to ignore, and this is not an incidental feature of biological motivation but close to its central design principle. Pain interrupts ongoing behavior in a way that milder signals do not. Fear recruits attention across the whole organism rather than remaining a localized signal. Hunger creates urgency that grows, rather than simply persisting, the longer it is ignored. Social rejection can dominate cognition for hours or days, crowding out competing concerns that might, on reflection, matter more. A system whose errors hurt has, in short, an extraordinarily effective mechanism for ensuring that at least some discrepancies acquire priority over the countless others competing for the organism's limited attention.

But this is not the only conceivable mechanism for achieving that same functional result, and it is worth taking seriously the possibility that it is not even the best one. If positive affect can carry sufficiently fine information, as Pearce's proposal claims it can, then an engineered organism could in principle respond vigorously to a difference between two good states without requiring either state to be bad, using intensity or texture of positive affect to do the interrupting and prioritizing work that pain does in biological organisms.

The inversion this suggests is striking, and it runs against the intuitive expectation that a world with less suffering would also, more or less automatically, be a world with less discrimination. As civilization moves north, it might become *more* sensitive to difference rather than less, not despite the disappearance of suffering but partly because of it, since the disappearance of the coarse, attention-dominating signal of pain leaves computational and attentional resources free to be redeployed toward finer discrimination among what remains.

A nervous system operating across extremely fine positive gradients could detect discrepancies too small for contemporary humans to notice at all, let alone act on. A social system populated by such agents might articulate distinctions that appear microscopic from the perspective of an age accustomed to famine, violence, disease, chronic pain, and involuntary death, distinctions that would strike a present-day observer as almost comically fastidious, in much the same way that a complaint about wine temperature can strike someone still drinking juice.

The disappearance of agony would therefore not imply the disappearance of

$$\text{things being worse than they could be.}$$

Indeed, increasingly sophisticated intelligence, freed from the coarse override of biological pain, may actually improve the detection of precisely such differences, extending the reach of discrimination further into regions of experience that presently register, if they register at all, only as vague and undifferentiated contentment.

Hence:

$$\boxed{\text{abolition of suffering} \;\not\Rightarrow\; \text{abolition of grievance.}}$$

This is not an objection to abolishing suffering, and it should not be read as one; nothing in the argument counsels against Pearce's project or any more modest version of it. It may instead be what successful abolition actually looks like, once the project is examined closely enough to distinguish what it eliminates from what it necessarily preserves. The engineering objective, on this reading, would not be to preserve agony because agony once carried useful information, which would be a perverse and self-defeating goal. It would be to preserve the information itself while changing the medium in which that information is represented, carrying forward the function of suffering, discrepancy detection, without carrying forward its phenomenological cost.

Or, more simply, in the phrase this essay keeps returning to:

> Keep the information. Lose the agony.

## 13 A Sentence That Doesn't Cash Out

The distinction between these four categories becomes harder to maintain the moment the argument approaches its own limit, and Pearce's writing supplies an unusually clean example of exactly where the difficulty lies. Pearce has written:

> "I would indeed be overjoyed if such suffering were to end—even if the price were to be no sentience at all."[3]

The surrounding position, read in full rather than in isolation, is considerably more qualified than this sentence alone suggests. Pearce also emphasizes the legal sanctity of life as a principle he wants enshrined regardless of his own deeper ethical commitments, explicitly distinguishes his practical abolitionist project from any simple commitment to strict negative utilitarianism, and treats the persistence of involuntary suffering itself as potentially relevant to catastrophic and existential risk, on the grounds that severe enough suffering can motivate people to act destructively toward others. The single sentence quoted above should therefore not be made to carry the entire weight of his position, and doing so would misrepresent an author who has, in the same breath, argued against the very outcome the sentence appears to endorse.

Nevertheless, its grammar is revealing on its own terms, independent of the surrounding qualifications, and it is worth examining that grammar directly. Suppose sentience disappears completely, so that

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

There is then no subject remaining to instantiate the state described by

$$H = \text{overjoyed}.$$

Strictly interpreted as a description of the resulting world, the sentence cannot cash out, because there is no post-abolition experiencer whose joy could constitute the evaluation the sentence attributes to that world. Joy is not a free-floating property of states of affairs; it is, on any plausible account, a property of a subject's relation to a state of affairs, and the sentence as literally construed removes the subject while retaining the property.

The most charitable reading available is temporal rather than metaphysical. The present speaker anticipates a possible future and presently experiences, or imagines experiencing, joy at the prospect that suffering has ended, a perfectly coherent psychological state that requires no subject in the imagined future at all, only a subject in the present doing the imagining. Thus

$$J_{\text{present}}\,(\text{future absence of suffering}) > 0$$

does not imply

$$J_{\text{sentience-free future}} > 0.$$

The latter expression has no subject to which it could refer, and treating it as though it did is precisely the kind of category slippage this essay is trying to guard against throughout.

This is an Epicurean inversion of sorts, running parallel to the ancient argument that death cannot be bad for the person who dies because there is no one left for it to be bad for. If no sentience exists, there is equally no one there for the absence of suffering to feel good to, just as there is no one there for it to feel bad to; the entire evaluative apparatus, positive and negative alike, requires a subject to operate at all.

The point of raising this is not to score a cheap verbal contradiction against abolitionism as a movement, and Pearce's own more careful formulations elsewhere make clear he does not actually endorse a sentience-free outcome as desirable. The point is to observe how easily the four distinctions from the preceding section collapse under conceptual pressure, even in the writing of an author unusually attentive to exactly these distinctions. A proposition about negative valence becomes,

almost without anyone noticing the transition, a proposition about evaluation. An evaluation becomes an imagined affective response. That response is then grammatically projected into a state in which, by the argument's own stipulation, there is no subject capable of possessing it, and the sentence still reads, on first encounter, as though it made sense.

If specialists discussing the outer limits of suffering abolition can slide between these coordinates under conceptual pressure, ordinary grievance-talk, conducted with far less care and under far greater emotional pressure, should be expected to do so even more readily and far more often. This matters directly to the North Setpoint, because reference-relative dissatisfaction can easily be mistaken for absolute suffering in exactly this way, while absolute positive valence can just as easily be mistaken for the complete absence of negative evaluation. The four categories are not interchangeable, however naturally the language available to describe them treats them as though they were.

## 14 The Bridge to *What Affect Permits*

The North Setpoint, as developed so far, concerns the formation of grievance: how a discrepancy between experience and expectation comes to be registered as negative at all. It does not, by itself, explain why a registered grievance becomes speech, confrontation, withdrawal, protest, accusation, criticism, or complaint, and a full account of observed complaint behavior needs a second, companion mechanism to close that gap.

The combined pipeline linking the two can be represented as

experienced conditions $\longrightarrow$ reference adaptation $\longrightarrow$ discrepancy $\longrightarrow$ grievance $\longrightarrow$ expression threshold $\longrightarrow$ c

The North Setpoint concerns primarily the left and middle portions of this chain. It asks how changing experiential conditions alter the scale on which discrepancies become salient in the first place, and it stops once a grievance has been formed, without making any further claim about what happens to that grievance next.

The companion account, *What Affect Permits*, concerns the later gate in the same pipeline. A grievance can exist without ever being expressed, held privately for reasons entirely separate from its severity, just as an expression can be facilitated by changes in social cost, inhibition, self-monitoring, perceived safety, anonymity, exhaustion, intoxication, or affective state, none of which need bear any direct relation to how the grievance was formed in the first place.

Schematically, let $a_i$ represent activation of a behavioral disposition toward expression, and $\theta_i$ its expression threshold. Then

$$b_i \text{ is expressed} \quad \Longleftrightarrow \quad a_i > \theta_i.$$

The present essay concerns primarily how something becomes worth negatively registering in the first place, the left-hand side of this pipeline. The companion essay concerns how something already registered becomes permissible, safe, or likely to express, the right-hand side. Thus complaint production, taken as a whole observable phenomenon, contains at least two independently moving parts:

$$\boxed{\text{threshold for grievance} \neq \text{threshold for expression.}}$$

Rather than leave this claim in purely verbal form, the companion essay's cleanest result can be imported directly rather than duplicated in prose. Let $g(t) = P(\text{grievance})$ denote the probability

that a given discrepancy is registered as a grievance at all, and let $e(t) = P(\text{expression} \mid \text{grievance})$ denote the conditional probability that a registered grievance crosses its expression threshold and becomes complaint. Observable complaint is then the product

$$c(t) = g(t)\, e(t).$$

Differentiating gives the identity that does most of the remaining work:

$$\boxed{\dot{c} = e\dot{g} + g\dot{e}.}$$

This single equation replaces what would otherwise take several paragraphs of case-by-case reasoning about who becomes sensitive while remaining silent, or freer to speak while suffering less. The North Setpoint, together with the mechanisms developed in the preceding sections, principally explains variation in $g(t)$, how a discrepancy comes to be registered as a grievance in the first place. *What Affect Permits* principally explains variation in $e(t)$, how a registered grievance comes to be voiced. Complaint statistics give direct access only to $c(t)$, their product, and the decomposition shows precisely why that access is not enough on its own to recover either factor. A rise in $c(t)$ is consistent with a rise in $g(t)$, a rise in $e(t)$, or both; a fall in $c(t)$ is consistent with a fall in either term, or with one term rising while the other falls faster still. Without independent evidence bearing on $g(t)$ and $e(t)$ separately, $c(t)$ alone cannot settle which has occurred, and mistaking a change in one for a change in the other is exactly the inferential error this bridge exists to block.

## 15 Against the Reactionary Reading

There is an obvious misuse of the North Setpoint argument, and it is important to name it explicitly rather than leave it as an implicit risk for a careless reader to discover on their own. If contemporary people complain about smaller harms than their ancestors endured, one might conclude that contemporary grievances are decadent, illegitimate, or evidence that suffering reduction has made people weak, unable to tolerate difficulties that earlier generations absorbed without comment. Nothing in the model developed here entails this conclusion, and the model can in fact be used to show precisely why the inference fails.

An event can simultaneously be

minor by an absolute historical comparison

and

large relative to the subject's adapted experiential range.

Both propositions can be true at once, describing the very same event from two different and equally legitimate vantage points, and neither cancels the other or renders it somehow less true. The fact that earlier populations endured cholera does not determine whether a contemporary grievance concerning humiliation is justified, because the two grievances are being measured against entirely different reference points formed under entirely different conditions, and there is no principled way to translate directly between them. The fact that someone today is protected from famine does not establish that betrayal cannot matter to them, any more than the fact that a wine drinker is protected from having to drink spoiled juice establishes that serving temperature cannot matter to them. Historical magnitude and present normative legitimacy are, in short, different questions, answerable by different kinds of evidence, and collapsing them into one question is the central error this section is written to block.

Indeed, increasing sensitivity to smaller discrepancies may be one entirely predictable consequence of moral and material progress, rather than a symptom of its failure. Once murder is rare, assault becomes more salient by comparison, occupying more of the remaining evaluative space. Once assault is rare, coercion becomes more salient in turn. Once overt coercion is rare, subtler forms of manipulation, barely registered as harms in a harsher environment, may become more salient still, each step following the same renormalization logic developed in earlier sections. There is no logical point at which improvement requires the evaluative system to stop discriminating, and expecting one to appear is, on this account, a category error about how bounded evaluative systems work.

This does not mean that every progressively smaller grievance is thereby correct, and reference-relative salience is emphatically not offered here as a moral proof of anything. A grievance can be intensely experienced and mistaken, felt with full sincerity while resting on a misunderstanding of the facts. It can be relatively enormous, occupying nearly the entire span of the person's current evaluative range, and absolutely trivial by any external standard. It can be absolutely serious and yet socially ignored, failing to register with anyone but the person experiencing it. The North Setpoint explains none of these cases away, and it is not designed to; its purpose is diagnostic rather than adjudicative. It warns against inferring the absolute magnitude of suffering directly from the intensity, frequency, or granularity of expressed grievance, without offering any shortcut for determining, in a given case, whether a particular grievance is in fact justified.

The reactionary mistake this section is written against can be stated compactly as

$$\text{smaller objective discrepancy} \Rightarrow \text{illegitimate grievance.}$$

The opposite mistake, equally worth naming and equally unsupported by the argument, is

$$\text{intense grievance} \Rightarrow \text{large absolute harm.}$$

The theory developed here supports neither implication, and a mature evaluative framework, whether held by an individual or embedded in an institution's policy toward complaints, should be capable of holding absolute and relative descriptions of the very same event simultaneously, without collapsing one into the other for the sake of a simpler verdict.

## 16 Not Every Northward Movement Is Progress

The reactionary reading just addressed establishes that relative salience does not establish moral illegitimacy: a grievance can be small in absolute terms and still be justified, or large in absolute terms and still be mistaken, and the theory offers no shortcut past that judgment. But guarding against that misreading leaves an obvious converse question unaddressed, one this essay has deferred until the machinery needed to answer it carefully was fully in place. If the reference point continually moves north, as the preceding sections have argued it often does and as the next section will argue it may do without any final terminus at all, how is genuine refinement of the evaluative scale to be distinguished from pathological calibration of it?

The generalized reference operator introduced earlier makes the question precise rather than merely rhetorical. Recall

$$N_i(t) = \alpha_i N_i^{\text{self}}(t) + \beta_i N_i^{\text{social}}(t) + \gamma_i N_i^{\text{cf}}(t).$$

Nothing in this construction guarantees that $N_i^{\text{social}}(t)$ is representative of any population worth being measured against, rather than, say, an unrepresentative slice of visible peers selected by

algorithmic curation, geographic accident, or status-driven self-presentation. Nothing in the construction guarantees that $N_i^{\text{cf}}(t)$ tracks states that are actually attainable, rather than states that merely feel attainable because they have been vividly described, persuasively marketed, or aspirationally performed by others with an interest in raising exactly this reference point. A setpoint can rise for excellent reasons, because better possibilities have genuinely become reachable and the evaluative system has correctly registered that fact, or it can rise for much worse ones, through misinformation, prestige competition, distorted or unrepresentative comparison populations, counterfactuals that were never really available, or outright adversarial manipulation by parties who profit from a population's chronic sense of falling short.

The essay's formal apparatus already contains the resources needed to mark this difference, even though it has not, until now, been given its own name. Distinguish a descriptive reference point from a warranted one:

$$N_i^{\text{descriptive}} \neq N_i^{\text{warranted}}.$$

The descriptive reference point is simply whatever $\mathcal{N}(\mathcal{H}_i, \mathcal{S}_i, \mathcal{C}_i)$ happens to output given the actual, causally upstream inputs feeding it, however those inputs came to be what they are. The warranted reference point is whatever the same operator would output given inputs that were accurate, representative, and genuinely attainable, filtered through whatever epistemic and normative standards the evaluator would endorse on reflection. The two coincide under favorable conditions, when social comparison is representative and counterfactuals are real, and they can diverge sharply under unfavorable ones, when comparison populations are curated for effect or counterfactuals are illusions maintained for someone else's benefit.

This distinction matters because it forecloses an objection that would otherwise remain fully available to a skeptical reader of everything argued so far: if the model can explain arbitrarily escalating expectations as a natural consequence of its own mechanics, does the model thereby vindicate them, treating every rising standard as evidence of genuine progress simply because a rising standard is what the mathematics predicts? The answer is plainly no, and the model was never built to imply otherwise. The theory explains the formation and salience of a reference point; it does not, by itself, certify that reference point as justified. This is close to a claim the essay has already made once, in the discussion of the reactionary reading, where the theory was described as diagnostic rather than adjudicative. What this section adds is a mathematical home for that qualification, so that the gap between description and warrant is not merely an informal caveat appended to protect the theory from misuse, but a distinction built into the same operator that generates $N_i(t)$ in the first place, and therefore a question that can, in principle, be investigated with the same rigor as everything else in this essay: not whether the setpoint has moved, which the model can already tell us, but whether it should have.

A concrete illustration of this divergence, running in the opposite direction from the escalating-standards case that most naturally comes to mind, appears in the later chapters of the ancient apocalyptic text attributed to Enoch. There the rhetoric turns repeatedly against people who have accumulated wealth through injustice and who mistake their own present comfort for evidence that no reckoning is coming. On the reading offered here, their condition is not a failure of the renormalization mechanism itself but a case in which $N_i^{\text{social}}(t)$ has been captured: their comparison population consists disproportionately of other people who benefited from the same injustice, so the social term feeds back a falsely reassuring signal, while attainability and justificatory weight, $A_i$ and $J_i$, are suppressed by the same self-interest that produced the wealth in the first place. Their reference point holds still, or even falls, exactly where an honest accounting would have required it to rise, since increased material security ought to sharpen rather than dull the capacity

to notice injustice once bare survival is no longer the organizing concern. This is the mirror image of the escalating case discussed above: an unwarranted failure to rise rather than an unwarranted rise in demandingness, $N_i^{\text{descriptive}}$ lagging $N_i^{\text{warranted}}$ instead of overtaking it, produced by exactly the same mechanism, a distorted or self-serving comparison population, running in the opposite direction.

A related but distinct failure appears earlier in the same corpus, in its account of a class of heavenly beings who instruct humanity in weapon-making, astrology, and other arts before human institutions possess anything resembling the capacity to absorb them safely. That case, on close inspection, is not a story about a reference point moving at all; it concerns capability introduced into a receiving structure faster than the structure can assimilate it, a hazard belonging to questions of transmission and premature disclosure rather than to the renormalization mechanism this essay develops. The two ancient cases are useful together mainly for showing that a single literary source can index two structurally distinct hazards, one a distortion of the reference point of exactly the kind this section formalizes, the other a mismatch between generation and reception that lies outside this essay's scope.

## 17 The North Has No Final Latitude

Pearce's Far North gives this essay a direction of travel, a north toward which the reference point moves, but it does not by itself settle a question that becomes unavoidable once the direction has been granted: is there a final latitude, a point past which the reference point stops moving because there is nowhere further north left to go, or does the ascent continue without any terminal state at all?

Two possibilities are worth distinguishing formally. The first is bounded convergence, in which the reference point approaches some fixed ceiling as time proceeds,

$$N_t \to N_\infty,$$

with $N_\infty$ representing either a biological ceiling on attainable hedonic architecture, an informational ceiling on how finely any physically realizable system could discriminate between states, or some other hard limit built into the structure of the problem itself. The second is unbounded ascent, in which the reference point continues rising indefinitely,

$$\frac{dN_t}{dt} > 0 \qquad \text{indefinitely},$$

with no value of $N_\infty$ ever being approached, because there is always, in principle, further room to move.

The comparison-population extension developed earlier supplies a strong reason to suspect the second possibility is the more realistic one, at least for any sufficiently intelligent and self-modifying civilization. The counterfactual term $\gamma_i N_i^{\text{cf}}(t)$ tracks not merely what presently exists but what has become conceivable, and the space of conceivable arrangements does not obviously have a ceiling of its own. Every closed gap can open a view onto a further gap that was previously not even representable, because the very act of closing one discrepancy typically expands the imaginative and technical vocabulary available for identifying the next one. A civilization capable of engineering information-sensitive gradients of bliss is, almost by definition, a civilization whose capacity to conceive of still finer or still more expansive gradients is itself increasing over time. On this reading, the North Setpoint is not a thermostat with a maximum setting waiting to be reached. It is a

process without any final calibration, open-ended in the same way that the frontier of scientific or technological possibility has so far proven to be open-ended.

This conclusion can sound alarming on first encounter, and it is worth addressing that alarm directly rather than leaving it to fester unanswered in the background of an otherwise optimistic essay. If the reference point never converges, does progress terminate in a permanent, Sisyphean condition of endless dissatisfaction, in which no amount of improvement ever produces a stable resting state of contentment? The alarm rests on exactly the conflation this essay has been trying to dissolve from its opening pages onward: it treats the persistence of negative comparison and negative judgment, which this section grants may indeed be permanent, as though it were equivalent to the persistence of suffering, which the preceding sections have given independent reasons to think need not follow at all. An unbounded $N_t$ is entirely compatible with an unbounded $H_{\min}(t)$ rising alongside it, provided the pace and character of the two ascents keep every agent inside a corridor that remains, at every point along the way, positively valenced and by any earlier standard extraordinary. What is ruled out by an ever-receding $N_\infty$ is a terminal state in which nothing whatsoever remains to be discovered wrong. What is not ruled out, and what the earlier sections give reason to expect instead, is a civilization that is, at every point along that infinite ascent, better off than it was at the point before, without that ascent ever needing to reach an end in order to have been worth undertaking.

A cleaner criterion than convergence is therefore available for judging whether such a process should count as success rather than failure. Call a trajectory asymptotically sufficient if the gap between experience and reference does not widen over the relevant horizon, even while both terms diverge without limit:

$$\frac{d}{dt}\big(H_t - N_t\big) \;\geq\; 0.$$

A civilization satisfying this condition never needs its reference point to stop moving in order for its members to flourish. It needs only that its capacity for delivering better states keep pace with, or outrun, its capacity for imagining still better ones, indefinitely, one step at a time. There is no reason internal to the model to expect this race to have a finish line, and, on the argument developed across this essay, no reason to want one. The absence of a final latitude is not a defect in the North Setpoint. It is what keeping the information, without keeping the agony, actually looks like once it is run forward without an artificial stopping point inserted for the sake of narrative comfort.

## 18 Conclusion: Keep the Information, Lose the Agony

The North Setpoint begins with an apparent contradiction, one visible in ordinary experience long before it is given any formal treatment. Human beings can become extraordinarily safe while remaining extraordinarily capable of grievance. Severe harms can disappear from a population's lived experience while its arguments migrate toward distinctions that would have seemed microscopic, perhaps even unintelligible, under harsher conditions. Improvement therefore appears, from one angle, strangely unable to satisfy, producing populations that are by any external measure better off and yet no less occupied with the sense that something remains wrong.

The ordinary hedonic treadmill supplies only part of the explanation for this pattern, and treating it as the whole explanation leaves the pattern more puzzling than it needs to be. The stronger possibility developed across this essay is that the evaluative reference frame itself moves,

and moves in a way that is neither an illusion nor a failure but a direct consequence of the evaluative system doing its job under improving conditions. If

$$N_{t+1} > N_t,$$

then improvement changes not merely what is experienced but what counts, from the perspective of the experiencer, as a significant departure from expectation. If the lower tail rises at the same time, so that

$$H_{\min}(t+1) > H_{\min}(t),$$

then the remaining evaluative range can become increasingly fine-grained, packing the same discriminative capacity into a narrower absolute span and thereby making that capacity more, rather than less, sensitive to small differences within it. A molehill can become locally mountainous without becoming absolutely large, and both descriptions can be correct at once, evaluated from their respective vantage points.

Pearce's Far North pushes this structure to its conceptual limit, further than ordinary experience or the wine example alone could take it. Suppose suffering really is abolished, in the strong sense Pearce intends. Suppose the worst states available to future minds are better than the greatest experiences available to us today, by whatever margin engineering eventually makes possible. Intelligence would still require differences to matter to it, on pain of losing the very capacity for goal-directed behavior that made the improvement meaningful in the first place. Better and worse would still have to be discriminable if action were to remain sensitive to gradients of value at all, however far north those gradients had been relocated.

The resulting world could therefore satisfy

$$H_t > 0 \qquad \forall t$$

while continuing to contain

$$H_i < H_j.$$

Positive valence need not eliminate negative comparison. Negative comparison need not entail suffering. Negative judgment need not entail complaint. Complaint need not measure absolute harm. These four independence claims, each established in an earlier section, together allow a possibility that is obscured whenever suffering and grievance are carelessly treated as a single coordinate moving together in lockstep.

A civilization may become dramatically better while continuing to discover things wrong with itself, and this is not a paradox once the underlying mechanism is made explicit; it may become better *because* it continues to discover them, using exactly the discriminative capacity that renormalization keeps sharp to identify the next improvement worth making. The content of dissatisfaction can migrate north as the conditions that once made dissatisfaction agonizing disappear beneath it, leaving the discriminative function of dissatisfaction intact while stripping away its cost.

This is not a brief against suffering reduction, whether Pearce's radical abolitionism or the ordinary incremental work of medicine, politics, technology, and social improvement that occupies most of the space between here and any far north. It is an argument against using the mere persistence of grievance as evidence that suffering reduction has failed, an inference the argument of this essay shows to be invalid on its own terms. Nor should the argument be inverted into its mirror image: the existence of historical atrocities does not invalidate smaller contemporary grievances merely because the latter occupy a narrower absolute range than the former, since

narrower absolute range and illegitimacy are, as the previous section argued at length, entirely different properties of a grievance.

The important fact, restated one final time, is that two coordinates have been confused throughout the ordinary, pre-theoretical way of talking about suffering and complaint. One coordinate describes how bad an experience is in absolute terms, independent of anyone's expectations. The other describes how far an experience lies below what an adapted evaluative system has learned, often without conscious deliberation, to regard as normal, expected, preferable, or possible given everything it has recently encountered. Moving north changes both coordinates, but it does not change them in the same way or by the same amount, and confusing the two is what generates the apparent paradox this essay set out to dissolve.

The deepest implication of information-sensitive gradients of bliss is therefore not that future minds would cease to care about discrepancies, as a first and overly literal reading of abolitionism might suggest. It is almost the opposite. If suffering currently performs informational work in guiding behavior, then abolishing suffering successfully, rather than merely suppressing its symptoms, requires preserving that informational work in another form, since discarding it along with the suffering that currently carries it would produce not a utopia but an organism incapable of preference at all.

Future minds, on this reading, might detect more differences, not fewer, freed from the coarse override that biological pain currently imposes on finer discrimination. They might possess finer preferences, not weaker ones, extending the reach of evaluation into regions of experience that presently register only as vague and undifferentiated contentment. They might identify defects invisible to us while inhabiting states we would regard, without qualification, as ecstatic, precisely because their discriminative apparatus operates at a resolution ours cannot match.

What disappears, on this account, need not be the capacity to say

this could be better.

What disappears is only the requirement that saying so, or feeling it, must hurt. Suffering is not the same thing as the information suffering happens, in biological organisms, to carry, and the two can in principle be separated, however inseparable they have always seemed from the inside of a nervous system built by natural selection rather than by design.

Every quantity this essay has introduced along the way, the hedonic level, the reference point, the discrepancy, the relative severity, the grievance, and finally the expressed complaint, can now be assembled into a single generative system, read downward from conditions to observable behavior rather than encountered piecemeal as a sequence of separate arguments:

$$
\boxed{
\begin{aligned}
N_i(t) &= \mathcal{N}(\mathcal{H}_i, \mathcal{S}_i, \mathcal{C}_i), \\
\Delta_i(t) &= N_i(t) - H_i(t), \\
\rho_i(t) &= \frac{\Delta_i(t)}{N_i(t) - H_{\min}(t)}, \\
G_i(t) &= \Psi(\Delta_i, A_i, J_i), \\
c_i(t) &= P(G_i > 0)\, P(E_i = 1 \mid G_i > 0).
\end{aligned}
}
$$

Read from the top, conditions determine $H$; experienced history, observed comparison, and reachable possibility together determine $N$; $H$ evaluated against $N$ generates a discrepancy; the active experiential range converts that discrepancy into a relative severity; attainability and justificatory weight convert severity into grievance proper; and the companion theory of expression converts grievance into the complaint an outside observer can actually count. Nothing in this system was

introduced only to be discarded by a later section. Renormalization, comparison populations, the error budget, attainability, information without agony, the warrant behind a moving reference point, and the openness of the ascent itself are not competing explanations of the same phenomenon so much as successive stages through which one variable becomes the input to the next. The essay, read this way, is less a sequence of increasingly interesting observations about grievance than a single model progressively revealing its own machinery, each section supplying exactly the piece the next one needed in order to be stated precisely.

That is the North Setpoint, stated as compactly as the whole argument allows:

**Keep the information. Lose the agony.**

## References


[1] Pearce, D. *The Hedonistic Imperative*. Online philosophical manifesto and research programme, 1995. `https://www.hedweb.com/hedab.htm`

[2] Pearce, D. "Life in the Far North: An Information-Theoretic Perspective on Heaven." Posted by algekalipso, August 3, 2016. `https://qualiacomputing.com/2016/08/03/information-sensitive-gradients-of-bliss/`

[3] Pearce, D. Comment on "Profile: The Far Out Initiative." *Astral Codex Ten*, May 15, 2024. `https://www.astralcodexten.com/p/profile-the-far-out-initiative/comments`

[4] Brickman, P., D. Coates, and R. Janoff-Bulman. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology* 36, no. 8 (1978): 917–927.

[5] Diener, E., R. E. Lucas, and C. N. Scollon. "Beyond the Hedonic Treadmill: Revising the Adaptation Theory of Well-Being." *American Psychologist* 61, no. 4 (2006): 305–314.

[6] Kahneman, D., E. Diener, and N. Schwarz, eds. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 1999.

[7] Helson, H. *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior*. New York: Harper & Row, 1964.

[8] Parducci, A. *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 1995.

[9] Schwarz, N., and G. L. Clore. "Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology* 45, no. 3 (1983): 513–523.

[10] Frederick, S., and G. Loewenstein. "Hedonic Adaptation." In *Well-Being: The Foundations of Hedonic Psychology*, edited by D. Kahneman, E. Diener, and N. Schwarz, 302–329. New York: Russell Sage Foundation, 1999.