Optimization and Thermodynamics

From convergent attractors to algorithmic entropy: the thermodynamic cost of optimization for embedded agents

Version 2 (revised exposition)

Daniel C Satya Benson
Williams College

June 2026

Abstract

A useful characterization of powerful agents is that they reliably steer the world into a narrow region of outcome space, a region that would be extremely unlikely to arise under any random process. These notes develop this picture from first principles, pursuing three successive aims: to make the notion of "steering into a narrow region" mathematically precise, to establish the thermodynamic constraints that physics imposes on any process realizing it, and to resolve a conceptual gap that the classical formalism leaves open. We first characterize an optimizer behaviorally, as the *cause* of a convergent attractor: a physical entity whose presence drives the system it acts upon, from a broad range of initial conditions and despite perturbations, into a narrow target set, so that conditioning on the optimizer collapses an observer's uncertainty about the final state that system reaches. We argue that even an observer concerned solely with prediction possesses an objective, information-theoretic reason to single out such optimizers, establishing optimization as an observer-independent feature of the world rather than a stance projected onto it; translated into information theory, optimization is local entropy reduction in the system being acted upon. We then provide thermodynamic foundations for the analysis of such processes: the reversibility of microscopic dynamics, together with the second law of thermodynamics, forbids global entropy reduction, so an embedded optimizer must compensate for every local reduction within a consistent global accounting. We derive the second law from reversibility in a minimal Markov-chain setting, classify the three ways in which an agent can reduce the entropy of a subsystem while respecting information conservation, and quantify the third of these channels through the Touchette–Lloyd theorem (entropy reduction beyond a "blind" baseline must be paid for in mutual information between the agent and the environment), developed in an appendix. A self-contained toy thermodynamics built out of biased coins (Wentworth's generalized heat engine), in which every step of the requisite "bookkeeping" is elementary, is developed in a separate appendix. Finally, we confront a conceptual gap: the standard Gibbs–Shannon entropy is defined relative to a subjective probability distribution that the formalism

treats as exogenous, making "capacity to optimize" observer-dependent and leaving the classical tools inapplicable to the far-from-equilibrium, information-bearing states of which agents are composed. Algorithmic thermodynamics, due to Ebtekar and Hutter, replaces the subjective distribution with Kolmogorov complexity, yielding a second law that applies to individual physical states. Under this reframing, an embedded agent's probabilistic knowledge becomes an endogenous, physically encoded quantity—the algorithmic mutual information between the agent's memory and the environment—which is precisely the resource the agent expends when it optimizes. We thus establish that agents with greater knowledge of the world possess correspondingly greater capacity to act upon it, with the exchange rate between the two set by thermodynamics.

Contents

1 Introduction: the role of thermodynamics in agent foundations

Agent foundations may be understood as the search for *robust concepts* (sometimes called "true names") for notions such as optimization, goals, world models, and embeddedness, where a robust concept is one that retains its meaning under extreme

optimization pressure and for agents far more capable than any yet constructed. These notes pursue the true name of *optimization* along the descriptive route: rather than beginning inside an idealized rational agent, equipped with beliefs, preferences, and an expected-utility criterion, we begin from the physical world itself. This raises fundamental questions: What kind of process can reliably steer the world into a narrow target? And what thermodynamic cost does physics impose on such a process?

Thermodynamics supplies the answer to the second of these questions, for two reasons that we state at the outset and that the remainder of these notes develops systematically.

Optimization as Local Entropy Reduction. An optimizer concentrates probability mass: out of the many configurations the world could occupy, it funnels a broad set of possibilities down to a few. Concentrating probability mass, however, is precisely the reduction of uncertainty, and the reduction of uncertainty is the reduction of entropy (Section 2 makes this correspondence precise). This identification is not an analogy: it brings the full machinery of entropy (the second law, fluctuation theorems, the cost of erasing a bit) to bear directly on optimization, and the modern form of that machinery, *stochastic thermodynamics*, holds arbitrarily far from equilibrium. Consequently, physics constrains the shape of any optimizer that could actually be built.

The Physical Currency of Knowledge for Embedded Agents. We distinguish two idealized pictures of agency. A *dualistic* agent sits outside the world it acts upon, in the manner of a player at a console: its inputs and outputs are sharply delimited, and "what it knows" is a free parameter, a distribution supplied exogenously by the analyst. An *embedded* agent, in contrast, is part of the world it acts upon, so its knowledge is not free: it must be stored in some physical substrate, in a brain or a memory register composed of the same atoms as everything else. One of the principal conclusions of these notes is that once entropy is defined correctly (algorithmically, as a property of an individual state rather than of a subjective ensemble), an agent's knowledge of its environment becomes an objective physical quantity—the mutual information between its memory and the environment—and that this quantity is exactly the budget available to be spent on optimization.

We close this introduction with a remark on the nature of the contribution and its intended audience. For readers with a background in physics, the value of these notes lies primarily in translation rather than in new theorems: recasting familiar thermodynamic facts in information-theoretic language ties them to agent-foundations notions such as world models, optimization power, and the limits on embedded agents. For readers without such a background, the notes are self-contained: every thermodynamic idea is defined here in information-theoretic terms, no statistical mechanics is assumed, and the only prerequisite is familiarity with elementary discrete probability.

Structure of the Notes. Section 2 develops the probabilistic and information-theoretic apparatus on which all later sections depend, introducing entropy, mutual information, and divergence together with the coding interpretations that give these quantities operational meaning. Section 3 then addresses the characterization of optimization: we characterize an optimizer behaviorally, as a physical entity that drives the

system it acts upon, from a broad range of initial conditions and robustly to perturbation, into a narrow target (a convergent attractor); we translate this characterization into entropy reduction; and we argue that any observer concerned with prediction possesses an objective, information-theoretic reason to attend to optimizers, independently of who is watching. Section 4 confronts this picture with the reversibility of microscopic physics and derives the second law of thermodynamics, with full proof, in a minimal coarse-grained setting. Building on this foundation, Section 5 classifies the three ways in which an embedded agent can reduce the entropy of a subsystem while respecting global information conservation: exporting it into the environment, absorbing it into memory through measurement, or expending pre-existing mutual information with the subsystem. Section 6 exposes a conceptual problem at the heart of the standard formalism, namely that entropy, as conventionally defined, depends on a subjective probability distribution supplied from outside the physics. Motivated by this limitation, Section 7 develops the resolution, algorithmic thermodynamics: Kolmogorov complexity as entropy, the algorithmic second law, and an exact analysis of Maxwell's demon, with emphasis throughout on how an objective notion of entropy dissolves the subjectivity problem. Section 8 synthesizes these results into an account of embedded agency in which knowledge figures as an endogenous physical resource, and Section 9 collects the principal conclusions. Appendix A develops, in parallel to the main line of argument, a self-contained toy thermodynamics built out of biased coins (Wentworth's generalized heat engine), in which the analogues of heat, work, energy conservation, and the impossibility of perpetual motion can all be verified by hand; the main text refers to it wherever a concrete blind-policy example is instructive. Finally, Appendix B develops the Touchette–Lloyd theorem, which makes the third channel quantitative: the entropy reduction an agent can achieve beyond a blind baseline is bounded by the mutual information between its action and the environment, with fully worked examples.

2 Theoretical foundations: probability, entropy, and information

This section develops the probabilistic and information-theoretic apparatus on which the remainder of these notes relies. Readers already acquainted with Shannon entropy and mutual information may treat this section as reference material; however, we draw attention to the two components on which the later development relies most heavily: the coding interpretation of Section 2.3, which recurs throughout these notes, and the log-sum inequality (Lemma 2.9), which underpins our proof of the second law.

2.1 Random variables and notation

Throughout, capital letters X, Y, Z denote random variables taking values in countable (finite or countably infinite) sets, and lowercase letters x, y, z denote particular values. We write $\mu(x)$ for the probability that $X = x$ when the distribution is called μ, and we refer interchangeably to "the distribution of X" or "the ensemble" when speaking of μ.

A joint random variable (X, Y) has a joint distribution, and the conditional probability of y given x is written $P(y \mid x)$. All logarithms in these notes are base 2, so that information is measured in bits; the natural-log versions of every formula differ only by a factor of $\ln 2$.

2.2 Entropy

With this notation in place, we begin by introducing the central quantity of information theory, which measures the uncertainty associated with a random variable.

Definition 2.1 (Shannon Entropy)**.** The *Shannon entropy* of a random variable X with distribution μ is

$$H(X) := \sum_{x} \mu(x) \log \frac{1}{\mu(x)},$$

with the convention that terms with $\mu(x) = 0$ contribute zero. We also write $H(\mu)$ for the same quantity when we wish to emphasize the distribution rather than the variable.

Entropy quantifies uncertainty, namely the amount that an observer does not yet know about the value of X before observing it. We present several examples to calibrate the scale.

Example 2.2 (Calibration of the Entropy Scale)**.** A fair coin has entropy $\frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1$ bit, while a deterministic variable (one whose value is certain) has entropy 0 bits. A uniform distribution over all binary strings of length n has entropy n bits, and n bits is the maximum possible entropy for a variable with 2^n possible values, reflecting the fact that uniform distributions are the most uncertain ones. A *biased* coin that lands heads with probability p has entropy $h(p) := p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$, and two values of this function play a role in later sections of these notes: $h(0.2) \approx 0.72$ bits and $h(0.1) \approx 0.47$ bits. A coin biased toward one outcome is more predictable than a fair coin and consequently carries less than one bit of uncertainty.

2.3 The coding interpretation

Of the available interpretations of entropy, the one on which we rely most heavily is entropy as data compression. Suppose that the value of X must be transmitted to a colleague in binary, and that the objective is to minimize the average number of bits sent. The governing principle is to assign short codewords to the likely values and long codewords to the unlikely values, thereby reducing the expected description length by exploiting the structure of the distribution. Shannon's source coding theorem renders this principle precise: there exists a code (a prefix-free assignment of binary strings to

outcomes) whose codeword for outcome x has length essentially $\log \frac{1}{\mu(x)}$ bits, and no code can achieve a smaller average length than the entropy. Consequently,

$$H(X) \;=\; \text{the minimum average number of bits needed to describe a sample of } X.$$

Example 2.3 (A Small Optimal Code)**.** Let X take four values with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$, and assign the codewords `0`, `10`, `110`, `111`. Each codeword for an outcome of probability $2^{-\ell}$ has length exactly $\ell = \log \frac{1}{\mu(x)}$, and the average length is $\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1.75$ bits, which equals $H(X)$ exactly, confirming that the entropy is attained by a code whose codeword lengths match the ideal lengths $\log \frac{1}{\mu(x)}$.

We now assign a name to the quantity inside this expectation, since the central argument of Section 6 turns on the distinction between the individual quantity and its average.

Definition 2.4 (Shannon Codelength, Also Called Stochastic Entropy or Surprisal)**.** For an individual outcome x under distribution μ, the *Shannon codelength* is

$$\hat{H}(x, \mu) \;:=\; \log \frac{1}{\mu(x)}.$$

This quantity is the length of the codeword assigned to x under the code optimized for μ, and the entropy is its mean: $H(\mu) = \left\langle \hat{H}(X, \mu) \right\rangle_{X \sim \mu}$.

We emphasize the dependency structure of this definition. The codelength $\hat{H}(x, \mu)$ is a property of the *pair* consisting of an outcome and a distribution—the same physical outcome x receives a short codelength under a distribution that anticipated it and a long codelength under a distribution that did not. At this stage, no notion of "the entropy of an individual outcome x" exists on its own, in the absence of a distribution against which the outcome is read. This observation becomes the crux of Section 6.

2.4 Joint and conditional entropy, and mutual information

Having interpreted entropy as an optimal description length, we now extend these notions to pairs of random variables, characterizing both the uncertainty that remains in one variable once another is known and the information that the two variables share.

Definition 2.5 (Conditional Entropy)**.** For jointly distributed X, Y, the *conditional entropy* of X given Y is

$$H(X \mid Y) \;:=\; H(X, Y) - H(Y),$$

where $H(X, Y)$ is the entropy of the pair. Equivalently, $H(X \mid Y)$ is the average,

taken over values y, of the entropy of the conditional distribution of X given $Y = y$; it measures the uncertainty about X that remains once Y is known.

The rearranged identity $H(X, Y) = H(Y) + H(X \mid Y)$ is called the *chain rule*, and it admits a transparent coding interpretation: to describe the pair, one first describes Y at an average cost of $H(Y)$ bits, and then describes X using a code adapted to the known value of Y at an average cost of $H(X \mid Y)$ bits.

Definition 2.6 (Mutual Information)**.** The *mutual information* between X and Y is

$$\begin{aligned} I(X;Y) \;&:=\; H(X) + H(Y) - H(X,Y) \\ &=\; H(X) - H(X \mid Y) \;=\; H(Y) - H(Y \mid X). \end{aligned}$$

Mutual information is the average number of bits that knowledge of Y saves in describing X; by the symmetry of the definition, it is equally the number of bits that knowledge of X saves in describing Y. It vanishes exactly when X and Y are independent, and it is never negative, a fact we prove below. Mutual information serves as the standard measure of "how much one part of the world knows about another", an informal phrase that recurs many times in these notes.

Example 2.7 (Calibration of Mutual Information)**.** Let $X = (B_1, B_2)$ consist of two independent fair bits, and let $Y = B_1$ be the first of them. Then $H(X) = 2$, and once Y is known only the second bit remains uncertain, so that $H(X \mid Y) = 1$ and $I(X;Y) = 1$ bit: Y contains exactly one bit of information about X. If instead Y were an independent coin flip, $I(X;Y)$ would equal 0; if Y were a full copy of X, $I(X;Y)$ would equal 2 bits, the whole entropy of X, confirming that in this example the mutual information varies from zero under independence to the full entropy of X when Y determines X.

2.5 Divergence and two fundamental inequalities

Having quantified the information shared between two variables, we now introduce a measure of the discrepancy between two distributions, from which the inequalities underlying our subsequent analysis follow.

Definition 2.8 (Kullback–Leibler Divergence)**.** For two distributions μ, ν on the same countable set,

$$D(\mu \,\|\, \nu) \;:=\; \sum_{x} \mu(x) \log \frac{\mu(x)}{\nu(x)}.$$

In coding terms, $D(\mu\|\nu)$ quantifies the cost of holding a mistaken model of the source: if the true distribution is μ but compression is performed with the code opti-

mized for ν, the average expenditure is $H(\mu) + D(\mu\|\nu)$ bits rather than $H(\mu)$. The definition remains meaningful, and we will make use of it, even when the reference ν is an unnormalized nonnegative measure rather than a probability distribution.

Both of the inequalities required for our analysis follow from a single elementary property of the convex function $t \mapsto t \log t$.

Lemma 2.9 (Log-Sum Inequality)**.** *Let $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ be nonnegative reals with finite sums $a := \sum_i a_i$ and $b := \sum_i b_i > 0$. Then*

$$\sum_i a_i \log \frac{a_i}{b_i} \;\geq\; a \log \frac{a}{b},$$

with the conventions $0 \log \frac{0}{b} = 0$ and $a \log \frac{a}{0} = +\infty$ for $a > 0$.

Proof. The function $\varphi(t) = t \log t$ is convex on $[0, \infty)$. We apply Jensen's inequality (which states that for convex φ, the average of φ is at least φ of the average) with weights $w_i = b_i/b$ and points $t_i = a_i/b_i$:

$$\sum_i \frac{b_i}{b} \varphi\Big(\frac{a_i}{b_i}\Big) \;\geq\; \varphi\Big(\sum_i \frac{b_i}{b} \cdot \frac{a_i}{b_i}\Big) \;=\; \varphi\Big(\frac{a}{b}\Big).$$

Multiplying both sides by b and simplifying $b_i \cdot \frac{a_i}{b_i} \log \frac{a_i}{b_i} = a_i \log \frac{a_i}{b_i}$ yields the claim. □

Taking the a_i and b_i to be the probabilities of two distributions (so that $a = b = 1$) yields *Gibbs' inequality*: $D(\mu\|\nu) \geq 0$, with equality only when $\mu = \nu$. Nonnegativity of mutual information follows because $I(X;Y) = D\big(\text{joint distribution} \,\|\, \text{product of marginals}\big) \geq 0$. The second consequence, which we prove at its point of use in Section 4, is the *data processing inequality* for divergence: passing two distributions through the same noisy channel can only bring them closer together.

Having established these information-theoretic foundations, we now turn to the principal subject of these notes.

3 Characterizing optimization

3.1 Two notions of optimization and their relationship

The term "optimization" carries two distinct established meanings, and the relationship between them constitutes the central question of this section. In computer science, an optimization algorithm is a program that outputs the solution, or an approximation thereof, to an optimization problem: an objective function to be minimized over some feasible region. A program that returns $x \approx \sqrt{2}$, the minimizer of $y = (x^2 - 2)^2$ on the real line, is optimizing in this sense. In operations research and engineering, optimization denotes something physically more substantial: the reworking of a process or an artifact so that it better serves a purpose, as when a factory is reorganized

to produce more nails per unit cost. These two activities, optimizing a number inside a computer and optimizing a factory, are evidently related, yet the precise nature of their connection is far from obvious. Two questions therefore present themselves: What exactly is the relation between these two senses of optimization? What must a physical process accomplish in order to qualify as optimization? To address these questions systematically, we begin by distinguishing the roles involved in any optimization process.

3.2 Optimizers and convergent attractors

Our characterization rests on a distinction between two roles that must be kept separate. The first is the *system being optimized*: some part of the world whose configuration is of interest, and which, in the absence of intervention, could settle into a wide variety of configurations. The second is the *optimizer*: a separate physical entity whose presence drives that system reliably toward a narrow target. The optimizer is neither the target nor the funneling itself; it is the *cause* of the funneling. This idea admits the following crisp characterization: an optimizer is a physical entity such that conditioning on its presence collapses an observer's uncertainty about the final configuration of the optimized system.

The example of a courier delivering a package to a fixed address makes these two roles concrete. The package is the system being optimized, since its location could in principle be almost anywhere, while the courier plays the role of the optimizer. The package may begin at an arbitrary location in the city and be subjected to road closures and traffic perturbations along the way; the courier identifies a route around each obstacle, ensuring arrival at the designated address. Left to its own dynamics, the package exhibits no tendency toward any particular destination; it is the presence of the courier that drives it to the address. In the language of dynamical systems, the courier renders the address a *convergent attractor* for the package: a configuration toward which the package is pulled from far away, and to which it returns after being displaced. The attractor is a fact about the package's behavior; the courier is the entity that creates it.

Two features of this picture carry the entire analytical weight of our characterization: a *broad range of initial conditions* and *robustness to perturbation*. Regarding the first, the optimizer does not require the system to begin in any special configuration, since a wide variety of distinct starting points are all driven to the same destination. An observer who knows only that the package began somewhere in the city is highly uncertain about its initial location, yet upon learning that a courier is delivering the package, the same observer becomes confident about its final location. Regarding the second, the convergence survives disturbances applied while the process is underway. This second feature is what separates a genuine optimizer from a coincidence. A satellite coasting in orbit is subject to no steering influence; once perturbed, it drifts onto a new trajectory indefinitely, possessing no home configuration to which it returns. The courier, when blocked by a closed road, selects an alternative route and still arrives. The entire difference lies in whether some entity is present to supply the restoring

tendency.

One omission from this characterization matters for everything that follows: designating an entity as an optimizer asserts nothing about goals, preferences, or representations stored in a mind. A courier, a feedback circuit, a chess engine driving a game toward checkmate, a growing organism, a running optimization algorithm, and a team building a house are optimizers in exactly the same sense, each being a physical entity whose presence drives some system into a narrow target and holds it there against perturbation. What unifies these examples is a structural fact about their effect on the world rather than any mental attribute. The following three subsections develop this picture into a quantitative measure, establish why even an observer with no stake in the target has reason to attend to optimizers, and argue that the property of being an optimizer is consequently an objective feature of the world.

3.3 Optimization as entropy reduction

The convergent-attractor picture translates directly into the information-theoretic language of Section 2, and this translation provides the bridge to physics.

We model the configuration of the optimized system as a random variable, writing X for its initial configuration and Y for its final configuration. Before any optimizer is taken into account, X ranges over a broad set, so an observer who knows only that the system begins somewhere in that set faces high uncertainty $H(X)$. We now introduce the optimizer and condition on its presence. Since the optimizer drives the system into the narrow target regardless of where the system began, the residual uncertainty about the final configuration, $H(Y)$, is small. The gap between these two quantities is the *entropy reduction*

$$\Delta H \;:=\; H(X) - H(Y),$$

the number of bits by which the optimizer has narrowed the system's possibilities. A broad range of initial conditions makes $H(X)$ large; robustness is what preserves the magnitude of $H(X)$ while the convergence continues to operate; a narrow target makes $H(Y)$ small. A large ΔH therefore corresponds to reliably reaching a target that the system's own unaided dynamics would essentially never attain, providing a precise formulation of the informal statement with which these notes opened: a powerful optimizer reliably produces outcomes that would be extremely unlikely under any random process.

Remark (Scope and Limitations of the Entropic Summary)

The summary statistic ΔH deliberately discards the *direction* of optimization: it records how narrow the achieved region is, not which region it is nor whether any agent desired it. This loss is smaller than it may initially appear. Wentworth has observed that any expected utility maximization problem can be decomposed into two parts: an entropy-minimization component, and a component that shapes the world's distribution to resemble one particular target distribution (in his formulation, "utility maximization is description length minimization").

The present notes concern the first component, which is where the laws of physics impose their constraints; the second component (the identity of the target, and the reasons for its selection) belongs to the theory of goals and preferences rather than to thermodynamics.

3.4 The predictive value of optimizers

We now present a reason to attend to optimizers that presupposes nothing about any agent's goals. Consider an agent whose sole objective is to *predict* the future configuration of some part of the world. The claim of this subsection is that, among all the features of the environment such an agent could study, optimizers are among the least costly to characterize.

We begin with the ordinary case, in which no optimizer is involved. To predict the final state Y of a physical system, one must in general know its initial state X in detail and propagate that knowledge forward through the dynamics. This procedure is costly in two respects. Specifying the initial condition costs approximately $H(X)$ bits, and for a system of any appreciable size $H(X)$ is enormous. Moreover, the resulting prediction degrades as the forecast horizon extends: chaotic dynamics amplify any error in the knowledge of X, so that a coarse measurement of X conveys progressively less information about Y over time. A substantial informational expenditure in the present thus purchases a forecast whose quality only deteriorates.

Suppose instead that an optimizer is acting on the system, driving it toward a target set T. Two changes occur simultaneously, both operating in the predictor's favor.

First, *knowledge of the initial condition is no longer required.* The optimizer drives a broad range of initial states to T, so uncertainty about X never propagates into uncertainty about Y. The $H(X)$ bits that would otherwise have to be acquired are simply irrelevant to the forecast.

Second, *specification of the target itself constitutes the forecast.* The laborious route to predicting where the system ends up is to determine its exact starting configuration and to propagate the dynamics forward step by step. For any large system this route is infeasible: the starting configuration is prohibitively large to specify, and chaos ensures that the slightest error in it grows until the forecast becomes uninformative. The optimizer, however, supplies a considerably more economical route: since it drives the system to the target regardless of the initial configuration, the predictor may discard the starting configuration entirely and simply state the system's destination, namely the target. A target is informationally inexpensive to specify, and robustly so, since the optimizer reaches it despite disturbances that no predictor could have tracked.

This constitutes a saving of effort rather than the acquisition of information without cost. The statement that the system terminates at the target itself constitutes the forecast; no additional knowledge has been generated. The point is only that knowing the destination toward which a process is driven is informationally inexpensive and stable, whereas reconstructing where the process started, the sole alternative route to the same forecast, requires a description that is vast in extent and fragile under error.

Optimizers are worth identifying because they permit the prediction of an outcome that would otherwise be almost entirely inaccessible.

The courier example renders the two routes concrete. To predict where the package ends up, one could attempt to track the courier's vehicle, every traffic light, and every other car on the road, and integrate the entire configuration forward: an intractable accumulation of contingencies governed by chaotic dynamics, in which a single mistimed traffic light alters every subsequent turn of the route. Alternatively, one could learn a single fact—that the courier is delivering to 14 Elm Street—and predict that the package ends at 14 Elm Street, having modeled none of the traffic. The second route expends almost nothing and predicts almost perfectly, demonstrating that knowledge of the optimizer and its target substitutes for detailed knowledge of the dynamics.

The robustness property sharpens this conclusion further: because the optimizer reaches T in spite of perturbations, the forecast that the system ends in T survives disturbances that the predictor could neither have foreseen nor measured. The predictor is thus spared not only the cost of the initial condition but also the cost of modeling the perturbations, since the optimizer absorbs both. For this reason alone, an agent concerned exclusively with prediction has reason to scan its environment for optimizers and to track the targets toward which they steer, since no other feature of the environment compresses the future into so short a description.

3.5 The observer-independence of optimization

The argument of Section 3.4 answers a standing objection to treating "optimization" or "intelligence" as a genuine feature of the world at all.

The Observer-Relativity Objection. A well-known position holds that intelligence, and optimization with it, is *observer-relative*: a system counts as intelligent only because it performs well on the tasks that *we* happen to care about, judged by the standards that *we* happen to hold, and under a different choice of tasks the same system would appear unremarkable. This position has a formal backbone in the No Free Lunch theorems, which establish that, averaged over all possible problems, no optimizer outperforms blind search; competence is then never absolute but always relative to some restricted class of problems, and the choice of class reflects the observer's interests. The position has a philosophical counterpart in Dennett's intentional stance, according to which describing a system as an agent with goals is not a discovery about the system but a *stance* that an observer adopts because it pays off predictively, so that whether a system "has goals" depends on the observer's modeling capacities and convenience. On either formulation, the claim that X is an optimizer appears less like a fact about X than like a relation between X and some particular onlooker.

The Thermodynamic Response to the Objection. We first identify the point of agreement with the intentional stance: our account likewise grounds the significance of optimizers in their predictive value (Section 3.4). The disagreement concerns whether that value is relative to a particular observer, and we contend that it is not. The

question of whether some entity is driving a given system into a narrow target, from a broad set of initial conditions and robustly to perturbation, is a question about the system and the entities acting upon it, and its answer makes no reference to anyone's preferences. The predictive leverage that a known optimizer provides—a short and accurate description of the state toward which the system is driven—is identical for every predictor regardless of what, if anything, that predictor wants. Consequently, no choice of tasks or standards is required in order to identify optimizers: the bare objective of prediction, which any embedded agent has reason to pursue whatever its goals, already suffices to pick them out. The reason to attend to optimizers is therefore observer-*independent*.

None of this contradicts the No Free Lunch theorems, which concern average competence across *all* problems and assert nothing about whether some physical entity is driving a given system into a narrow target. Nor does our account deny that an optimizer has a target: the identity of the target is itself a property of the system, namely the region into which its dynamics actually funnel, rather than a verdict imposed from outside. The observer-relativity with which the objection is concerned, namely the question of whose standards adjudicate success, never enters the analysis, because we are not grading the optimizer against any standards at all; we are observing that it possesses an attractor and using that fact, as any predictor could, to predict the world. Optimization is accordingly an objective feature of the world itself rather than a construct that an observer projects onto it.

We now collect the conceptual framework before turning to the physics. An optimizer is a physical entity that drives the system on which it acts into a convergent attractor (Section 3.2); the amount of optimization is the entropy the optimizer removes from that system (Section 3.3); and any predictor, whatever its goals, has an objective reason to track optimizers (Sections 3.4 to 3.5). What remains is to determine what physics permits. Optimization constitutes entropy reduction in a subsystem—that is, *local* entropy reduction—and the following section establishes that *global* entropy reduction is physically impossible. This tension, and the "bookkeeping" that resolves it, organizes the remainder of these notes.

4 Reversibility and the second law

4.1 The reversibility of microscopic physics

Having introduced the convergent-attractor picture of optimization, we now examine the physical constraints that any such process must respect, beginning with the observation that, at the most fundamental level of description available to us, physical systems never lose information. In classical mechanics, the complete microscopic state of a system, its *microstate*, consists of the position and momentum of every particle, corresponding to a single point in the system's *phase space*. The laws of motion carry each microstate along a unique trajectory, determining the future from the present and, equally, the past from the present. It follows that two distinct microstates can never evolve into the same state: if they did, running the laws backward from the shared

future state could not determine which of the two microstates to recover, and the dynamics admit no such ambiguity. Over any fixed interval, therefore, the evolution of an isolated system constitutes a *bijection* on microstates, an invertible pairing of initial states with final states. Liouville's theorem strengthens this conclusion: the evolution preserves not only distinctness but phase-space *volume*, allowing a region of microstates to be stretched and folded without limit while retaining its total volume. Quantum mechanics expresses the same principle in different notation: isolated evolution is unitary, hence invertible and volume-preserving. In either formulation, microscopic information is neither created nor destroyed.

4.2 The incompatibility of global funneling with reversibility

Given the reversibility of the microscopic dynamics established above, we now confront the convergent-attractor picture of optimization with this fundamental property. Taken literally at the microscopic level, the two are directly incompatible: funneling requires that many initial states converge to the same small set of target states, constituting a many-to-one map, whereas reversibility guarantees that the microscopic dynamics are one-to-one, rendering a perfect funnel microscopically impossible.

A second, closely related obstruction arises from the second law of thermodynamics itself, the principle that the entropy of an isolated system tends not to decrease. In the framework we are about to develop, the second law is not an additional postulate but a *consequence* of reversibility, and understanding this derivation illuminates both principles. The resolution of the apparent paradox—optimization manifestly occurs, yet physics forbids funneling—is provided by the central "bookkeeping" principle of these notes:

> *No physical process can reduce the entropy of an isolated system. Optimization is entropy reduction in a coarse-grained description of a subsystem, and the information displaced from the optimized subsystem must be accounted for elsewhere in the larger system.*

The remainder of this section renders the first sentence precise and establishes it rigorously. Section 5 and Appendix B address the second sentence, characterizing where the displaced information can reside and quantifying how much entropy reduction an agent can accomplish.

4.3 Coarse-graining and the emergence of probability

The preceding discussion prompts a natural question: if the microscopic dynamics are deterministic and information-preserving, from where does probability arise, and what accounts for the rise of entropy? The answer lies in the fact that no observer ever tracks the microstate. A gram of matter possesses on the order of 10^{23} coordinates, each specified to unbounded precision, a description that no observer, and no embedded agent, can maintain. What is tracked instead is a *coarse-grained* description: we partition the

phase space into discrete cells, each aggregating all microstates that cannot be distinguished at the chosen resolution (for instance, every particle's position and momentum specified to finitely many digits, or merely a small collection of macroscopic variables), and describe the system by identifying the cell it currently occupies.

The dynamics of cells, in contrast to the dynamics of microstates, are genuinely stochastic. The current cell does not determine the next one, because different microstates within the same cell flow to different destinations. The most complete description available is a transition probability $P(y, x)$, defined as the fraction of cell x, measured by phase-space volume, that arrives in cell y one step later. This randomness is not metaphysical in character; it reflects precisely the microscopic detail that the coarse description declines to resolve. Chaotic dynamics continually transport such detail upward across the resolution scale, so that the coarse trajectory resembles a sequence of random jumps even though the underlying flow remains deterministic.

One structural assumption underlies the whole of stochastic thermodynamics: the coarse trajectory constitutes a *Markov process*, meaning that the distribution of the next cell depends only on the current cell and not on the earlier history. The intuition is that a sufficiently good coarse-graining ensures that the current cell contains everything about the past that is relevant to the coarse future, with the discarded detail contributing independent noise at each step rather than persisting as hidden memory. Whether a given coarse-graining is genuinely Markovian remains a deep and open question, but two considerations provide reassurance. First, there exist exactly solvable models, the *multibaker maps*: deterministic, time-reversible, chaotic systems whose coarse-grainings provably reproduce arbitrary Markov chains, with all of the randomness residing in the initial condition. These models demonstrate that macroscopic stochasticity and irreversibility are fully compatible with microscopic determinism and reversibility. Second, the Markov property is precisely what renders everyday statistical reasoning valid, and its *failure backward in time* constitutes the arrow of time. A dropped glass shatters at a predictable moment, and its shards obey well-defined local statistics that are entirely independent of events occurring at a neighboring house. Under time reversal this description fails: retrodicting the moment at which the shattered glass was dropped cannot be accomplished through local statistics, and the most informative evidence available to an observer may be a conversation about the accident taking place next door. Forward evolution obeys memoryless local laws while backward evolution does not, and this asymmetry is exactly the content of the Markov property.

One further consequence of reversibility remains to be extracted before we establish the second law. Liouville's theorem renders phase-space volume a *stationary measure* of the chain: weighting each cell by its volume produces a weighting that a single step carries to itself. We take all cells to have equal volume (a *mesoscopic* coarse-graining, obtained by truncating every coordinate to a fixed precision). Stationarity of the uniform weighting then determines the structure of the transition probabilities: with π constant, the condition $\sum_x P(y, x)\, \pi(x) = \pi(y)$ reduces to $\sum_x P(y, x) = 1$ for every y, so that the probabilities flowing *into* each cell sum to one, exactly as those flowing *out of* each cell do. A matrix satisfying both properties is termed *doubly stochastic*. Double stochasticity is thus the characteristic imprint of reversibility on the

coarse dynamics: a doubly stochastic step can permute and disperse probability but can never concentrate it, since concentration would require packing several cells' worth of volume into fewer cells, which Liouville's theorem forbids.[1]

4.4 The second law of thermodynamics

We now establish the second law for the class of doubly stochastic coarse-grained dynamics identified above, namely that the entropy of the coarse-grained state cannot decrease under a single step of the chain.

Theorem 4.1 (Second Law for Doubly Stochastic Chains)**.** *Let P be a doubly stochastic transition matrix on a countable state space, let the random variable X (the coarse-grained state now) have distribution μ, and let Y (the state one step later) have distribution $P\mu$, where $(P\mu)(y) := \sum_x P(y,x)\,\mu(x)$. Then*

$$H(Y) \;\geq\; H(X).$$

Proof. We establish the theorem in two steps.

Step 1: contraction of divergence under a single Markov step. Let μ and ν be any two distributions (or nonnegative measures) on the state space. We claim the data processing inequality

$$D(P\mu \,\|\, P\nu) \;\leq\; D(\mu \,\|\, \nu).$$

To verify this claim, we fix an output state y and apply the log-sum inequality (Lemma 2.9) to the numbers $a_x = P(y,x)\mu(x)$ and $b_x = P(y,x)\nu(x)$, whose sums over x are $P\mu(y)$ and $P\nu(y)$ respectively:

$$\sum_x P(y,x)\mu(x) \log \frac{P(y,x)\mu(x)}{P(y,x)\nu(x)} \;\geq\; P\mu(y) \log \frac{P\mu(y)}{P\nu(y)}.$$

We now sum this inequality over all y. On the left-hand side, the factor $\log \frac{\mu(x)}{\nu(x)}$ does not depend on y, and $\sum_y P(y,x) = 1$ because P is stochastic, so the left-hand side totals $\sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)} = D(\mu\|\nu)$; the right-hand side totals $D(P\mu\|P\nu)$, which establishes the claim. The underlying intuition is precisely the coding-theoretic one: passing two information sources through the same noisy channel can only render them more difficult to distinguish.

Step 2: specialization of the reference measure to the uniform measure. Let $\sharp$ denote the counting measure, $\sharp(x) = 1$ for every x (an unnormalized uniform measure; Lemma 2.9 at no point required normalization). Double stochasticity of P

[1] The entire framework developed in these notes generalizes to coarse-grainings with unequal cell volumes: entropy is then measured *relative to* the stationary volume measure π, replacing $H(\mu)$ by $H_\pi(\mu) := \sum_x \mu(x) \log \frac{\pi(x)}{\mu(x)}$ and the codelength by $\hat{H}_\pi(x,\mu) := \log \frac{\pi(x)}{\mu(x)}$. This generalization is standard, but it is not required for any of our subsequent results; the assumption of equal cells simplifies the notation at no conceptual cost.

states precisely that $P\sharp = \sharp$, since $(P\sharp)(y) = \sum_x P(y,x) = 1$. Moreover, divergence from the counting measure is simply negative entropy:

$$D(\mu \,\|\, \sharp) \;=\; \sum_x \mu(x) \log \frac{\mu(x)}{1} \;=\; -H(\mu).$$

Applying Step 1 with $\nu = \sharp$ yields

$$-H(P\mu) \;=\; D(P\mu \,\|\, P\sharp) \;\leq\; D(\mu \,\|\, \sharp) \;=\; -H(\mu),$$

which rearranges to $H(P\mu) \geq H(\mu)$. □

We draw attention to the ingredients of the proof: the list is short, and every item is physical in origin. The argument invoked the Markov property, which permitted a single step to be written as a transition matrix, and double stochasticity, which followed from Liouville's theorem, that is, from the *reversibility* of the microscopic dynamics; nothing further was required. Consequently, macroscopic irreversibility, the rise of entropy, stands in no tension with microscopic reversibility; it follows from it, once we grant that observers track coarse cells rather than microstates. This is the precise content of the assertion that "the second law follows from the reversibility of physics".

4.5 The scope and limitations of the second law

Three clarifications delineate the scope of this result and prevent misreadings that become consequential in later sections.

First, Theorem 4.1 applies to an *isolated* system evolving autonomously: the entropy of the whole cannot decrease. The theorem asserts nothing against the entropy of a *subsystem* decreasing, and the entropy of subsystems decreases constantly; this is precisely what refrigerators, crystallization, house-builders, and every other optimizer accomplish. What the theorem imposes is a bookkeeping constraint on the whole: when the entropy of a subsystem drops, the global accounting must remain consistent, with at least as much entropy appearing elsewhere in the total system. The complete classification of where this "elsewhere" can be located is the subject of Section 5.

Second, the theorem as stated concerns the entropy of the *ensemble*, that is, of the probability distribution μ, and it holds on average and in distribution; individual trajectories may pass through improbable, low-codelength states. A sharper, trajectory-level version of the second law, equipped with explicit and very small bounds on the permitted fluctuations, becomes available in the algorithmic framework of Section 7.

Third, and most consequentially for the remainder of these notes, the theorem assumed a fixed dynamics P uninfluenced by any observer, together with an entropy defined relative to a given distribution μ. Agents complicate both assumptions. An agent that *observes* the system and conditions its behavior on its observations does not constitute a fixed P; the extent of the advantage this confers, and the cost at which it is purchased, is the subject of Appendix B. Furthermore, the apparently innocuous reliance on a given μ conceals a deep conceptual problem, which we address in Section 6. Before turning to these questions, however, we take up the bookkeeping

question directly: when an agent lowers the entropy of a subsystem, the question arises of where the displaced entropy can reside, and the following section provides the complete answer. (A self-contained toy model in which every step can be verified by hand is developed separately in Appendix A.)

5 Three types of optimization under information conservation

5.1 The bookkeeping problem

Equipped with the second law and the reversibility of the underlying dynamics, we now undertake the global accounting that governs optimization by an embedded agent. We decompose the universe into three parts: the agent A, the subsystem S that the agent seeks to optimize, and the remainder of the environment E. Optimization of S consists in reducing $H(S)$, funneling the subsystem from a large set of possible configurations toward a narrow target set. The second law (Theorem 4.1, applied to the isolated whole) establishes that the joint entropy of (A, S, E) cannot decrease, and the reversibility underlying it demands that information about the initial condition of S cannot simply vanish. This raises the central "bookkeeping" question: through which channels can an agent reduce $H(S)$ while the global accounting remains consistent? The analysis of Daniel C and Ebtekar identifies exactly three such channels, and the classification is exhaustive: entropy removed from S must be transferred into E, absorbed into A, or paid for in advance through pre-existing correlation. We develop each of the three channels in turn, culminating in a reinterpretation of Maxwell's demon in terms of this classification.

5.2 Type 1: transfer of entropy into the environment as waste heat

The first channel is the most familiar of the three: the agent arranges an interaction between S and E under which $H(S)$ decreases while $H(E)$ increases by at least the same amount, thereby transferring the subsystem's entropy into the environment. A refrigerator narrows the distribution over the thermal state of its interior while heating the kitchen behind it; a house-building team imposes order on lumber and nails while dissipating heat, exhaust, and scattered debris into the surroundings. In the coin world of Appendix A, this channel is realized by a conditional-swap circuit that moves uncertainty from the designated coins into other coins. The signature of Type 1 is that the environment absorbs the entropy, the agent functioning merely as an intermediary that arranges the transfer and whose own state need not change appreciably in the process.

5.3 Type 2: absorption of entropy into the agent's memory through measurement

The second channel is subtler: the agent reduces $H(S)$ by *measuring* S and storing the outcomes, so that the entropy of the subsystem is transferred into the agent's own memory. This mechanism constitutes the operating principle of one of the most celebrated thought experiments in thermodynamics, which we now examine in detail.

Maxwell's Demon. A room full of gas is divided in two by a partition, and the partition contains a small door that can be opened and closed at will. A minute intelligence (the "demon") observes the molecules: whenever a molecule approaches the door from the left, the demon opens it, and whenever one approaches from the right, the demon closes it. The gas thereby accumulates, molecule by molecule, on the right side of the room. The entropy of the gas has evidently decreased (we are far less uncertain about the locations of the molecules than before), yet no work appears to have been performed and no heat has been discharged anywhere, presenting an apparent violation of the second law.

The second law has not been violated; rather, the reversibility argument of Section 4.1 identifies exactly where the missing entropy resides. Consider the end of the process and select any particle that now occupies the right side. Two distinct histories are consistent with its presence there: either the particle began on the right and remained, or it began on the left and was admitted by the demon. These two histories produce *identical* final gas configurations, so the gas alone cannot distinguish between them. The microscopic dynamics of the total system are reversible, however, and reversibility demands that the past be reconstructible from the present. The only way the total system can remain reversible is if some other component differs between the two histories, and that component is the demon's memory: the record of its observations and door operations. For every bit of entropy the demon removes from the gas, at least one bit of entropy accumulates in the demon's memory. The reduction in the entropy of the gas is thus fully compensated by the growth of entropy in the demon's records, ensuring that no global decrease occurs. (We revisit this argument with full rigor, for individual states rather than ensembles, in Section 7.4.)

The signature of Type 2 is that the agent absorbs the entropy: measurement transfers uncertainty from the world into the measurer. The demon's memory is finite, however, so this channel is exhaustible: once the memory fills, the demon must either halt or clear it, and clearing it ultimately exports the stored entropy to the environment as Type-1 waste after all (the exact bookkeeping is provided in Section 7.4). Type 2 therefore functions as a finite buffer rather than as a sink of unbounded capacity, deferring the environmental cost rather than eliminating it.

5.4 Type 3: expenditure of pre-existing mutual information

The first two channels balance the accounts by increasing entropy elsewhere, whether in E or in A. The third channel is the most surprising of the three: it reduces $H(S)$ *with-*

out increasing entropy anywhere, consuming instead a correlation that already exists between the agent and the subsystem.

Suppose that at time t the agent's state and the subsystem's state share mutual information $I(S_t; A_t) > 0$: the agent already "knows something" about the subsystem, in the purely statistical sense of Definition 2.6. To keep the accounting transparent, we assume that the environment is independent of both, so that the joint entropy decomposes (by the definition of mutual information and the chain rule) as

$$H(S_t, A_t, E_t) \;=\; H(A_t) \;+\; H(S_t) \;-\; I(S_t; A_t) \;+\; H(E_t).$$

The agent now performs an operation with the following effects: it *erases the copy of the shared information that resides inside* S, using its own copy as the key, while leaving its own marginal state and the environment untouched. The operation transforms the subsystem so that

$$H(S_{t+1}) \;=\; H(S_t) - I(S_t; A_t), \qquad I(S_{t+1}; A_{t+1}) \;=\; 0,$$
$$H(A_{t+1}) = H(A_t), \qquad H(E_{t+1}) = H(E_t).$$

Summing the joint entropy after the operation yields

$$H(A_{t+1}) + H(S_{t+1}) + H(E_{t+1}) \;=\; H(A_t) + H(S_t) - I(S_t; A_t) + H(E_t),$$

which is *exactly* the joint entropy from before. The entropy of the subsystem has genuinely fallen by $I(S_t; A_t)$ bits, the total entropy is unchanged, no heat has been produced, and no memory has been filled. What has been consumed is the correlation itself: afterward $I(S_{t+1}; A_{t+1}) = 0$, and the operation cannot be repeated without first re-acquiring mutual information.

Example 5.1 (The Controlled-NOT as a Minimal Instance of Type 3)**.** We illustrate the mechanism through the smallest possible instance, which renders the bookkeeping fully transparent: let the subsystem hold a single uniformly random bit s, and let the agent hold a perfect copy $a = s$. Then $H(S) = 1$, $H(A) = 1$, $I(S; A) = 1$, and the joint entropy is $H(S, A) = 1$ bit (two equally likely joint states, 00 and 11). The agent now applies a *controlled-NOT*, which flips s if $a = 1$ and leaves s unchanged if $a = 0$. In both joint states the result sets s to 0: the subsystem becomes deterministic, with $H(S_{t+1}) = 0$, corresponding to a full bit of entropy reduction. The operation is reversible (a second application undoes it), involves no environment, and produces no heat. The joint entropy remains exactly 1 bit, now carried entirely by the agent's own marginal (a remains a uniform bit, uncorrelated with the now-deterministic s). The correlation has been spent: a further application of the operation accomplishes nothing, since $I(S_{t+1}; A_{t+1}) = 0$. This confirms that a Type-3 operation reduces the entropy of the subsystem by exactly the mutual information consumed, with the global accounting remaining consistent throughout.

5.5 A reinterpretation of Maxwell's demon, and a synthesis of the three channels

Having identified the three channels individually, we now return to Maxwell's demon, whose operation the Type-3 channel recasts in an illuminating way. The demon's measure-and-control behavior can be decomposed into two distinct steps: a *copy* step, in which the demon interacts with the gas to acquire mutual information with it (this is Type 2: the demon's memory absorbs entropy, or more precisely, becomes correlated with the gas), followed by a *control* step, in which the demon expends that mutual information to steer the gas into a narrower distribution (this is Type 3: the correlation is consumed, and the control step itself produces no further entropy anywhere). In this decomposition, measurement constitutes the purchase of correlation and steering its subsequent expenditure, which raises a quantitative question: how much entropy reduction can the control step purchase per bit of correlation acquired in the copy step? This is precisely the question answered by the *Touchette–Lloyd theorem* (Appendix B), which bounds the steering achievable in a Type-3 step by the mutual information available to be spent, bit for bit. (Section 7.4 subsequently derives an exact algorithmic version of the same bound.)

To summarize, an embedded agent can optimize a subsystem in exactly three ways:

1. **Type 1:** transfer of the subsystem's entropy into the environment as waste heat;
2. **Type 2:** absorption of the subsystem's entropy into the agent's own memory through measurement;
3. **Type 3:** expenditure of mutual information already shared with the subsystem, erasing the subsystem's copy of the correlation.

These three channels are not mutually exclusive but compose in practice: a realistic optimizer cycles among them, measuring (Type 2), steering (Type 3), and periodically clearing its memory into the environment (Type 1) in order to free capacity for the subsequent round.

The Touchette–Lloyd theorem, developed in Appendix B, prices this exchange at one bit of entropy reduction beyond the blind baseline per bit of mutual information between action and environment, and we draw on its statement in what follows.

6 Limitations of subjective entropy

Throughout the preceding sections we have treated the distribution μ, against which all of our entropies are defined, as simply given. In this section we subject this assumption to systematic examination, and we find that it fails precisely in the regime where agent foundations requires it to hold. We begin by recalling the role of entropy as a measure of optimization capacity, then examine three distinct failure modes of the ensemble picture, culminating in the case of the demon's memory, which we argue constitutes the central case for agent foundations. The critique developed here, together with its

resolution in the following section, follows the essay of Daniel C and Ebtekar and the foundational paper of Ebtekar and Hutter.

6.1 Entropy as a measure of optimization capacity

We begin by recalling the function that entropy serves: across physics and engineering, the role of entropy is to measure a system's capacity to do work, and the developments of Sections 4 to 5, together with the coin world of Appendix A, permit us to read "capacity to do work" as *capacity to optimize*: negentropy constitutes the fuel that funnels broad distributions into narrow ones.

We now apply this reading to the demon after an extended run of Type-2 optimization. Its memory comprises, say, m bits, some of which are now occupied by measurement records. The demon's *remaining* capacity to optimize the gas is given by its free memory, namely the total size minus the entropy already stored, with each further bit of optimization consuming another bit of free memory. This observation raises the question on which the entire analysis turns: *Is the demon's free memory an objective, physical quantity?* One would expect this quantity to be objective: whether the hardware can absorb another billion measurements is a fact about the demon's physical configuration, carrying direct engineering consequences (how much longer the demon can continue to refrigerate, and how much heat it must eventually dissipate to clear its records). It would be anomalous for the demon's actual ability to act on the gas to depend on the beliefs of some outside observer describing it, yet this is precisely what the standard formalism delivers.

6.2 The exogenous status of the reference distribution

We recall the shape of the standard definitions (Definition 2.4). The entropy of an *individual* state x is undefined; what is defined is the Shannon codelength $\hat{H}(x, \mu) = \log \frac{1}{\mu(x)}$ of a state *relative to a distribution* μ, together with its average, the Gibbs–Shannon entropy $H(\mu)$. The distribution μ enters as an input rather than an output: it encodes the knowledge of some observer of the system, and the formalism specifies neither whose knowledge is encoded, nor where it is stored, nor how it was obtained, nor what it cost. In our terms, the standard framework treats knowledge as *exogenous*.

For the classical equilibrium applications of thermodynamics this exogeneity is harmless, because the appropriate choice of μ is essentially forced. This harmlessness, however, rests on three conditions. We say that μ constitutes a *useful summary* of an actual physical state x when all three of the following hold:

1. **Simplicity:** μ admits a short description, such as "an ideal gas of N molecules at temperature 300 K and pressure 1 atm, at equilibrium". (Otherwise, the specification of μ itself introduces unaccounted-for information, a point to which we return below.)

2. **Concentration:** the codelength concentrates near its mean, so that $\hat{H}(X, \mu) \approx H(\mu)$ with high probability under $X \sim \mu$. For systems composed of many weakly

interacting parts this property follows from the law of large numbers, and it is precisely what permits a single number, "the entropy", to stand in for an entire distribution of codelengths.

3. **Typicality:** the actual state x constitutes an unremarkable sample from μ, falling within the high-probability set where the previous two conditions apply.

When all three conditions hold, $H(\mu) \approx \hat{H}(x, \mu)$, every reasonable notion of entropy agrees, and thermodynamics may be conducted with macrovariables alone, without ever inquiring where μ is stored. Equilibrium statistical mechanics lies entirely within this regime, and the classical apparatus of Clausius and Boltzmann, with temperature, pressure, and chemical composition as the privileged macrovariables, is well adapted to it.

6.3 Three failure modes of the ensemble formalism

Far from equilibrium, and above all for *information-bearing* states, each of the three conditions identified above can fail independently, and each failure breaks the formalism in its own instructive way. We examine the three failure modes in turn.

Failure of Simplicity: Knowledge Without Informational Accounting. Let μ be a point mass on one particular, intricate configuration x, corresponding to exact knowledge of the microstate. Then $H(\mu) = 0$, and the formalism declares the system fully known and hence fully available for work extraction: an agent with exact knowledge of the state of a gas can in principle extract all of its energy as work, employing a transformation tailored to that state. The tailored transformation, however, together with the knowledge it embodies, must itself have been specified; the description of μ is exactly as intricate as x itself, and the formalism assigns it a cost of zero. The Gibbs–Shannon account thus permits arbitrarily detailed knowledge as an exogenous resource obtained at no cost. Physically, any agent in possession of such knowledge must encode it in a memory, and acquiring, storing, and clearing that memory carry costs to which the second law is sensitive, as the analysis of the demon has already demonstrated. The ensemble formalism provides no mechanism for accounting for these costs.

Failure of Concentration: Averages Representative of No Individual State. Consider a robot that flips a hidden fair coin and, depending on the outcome, either fully drains its battery or leaves it fully charged. Under the resulting equiprobable mixture μ over the two outcomes, the Gibbs–Shannon entropy of the battery together with its surroundings lies exactly halfway between the entropy of the charged state and that of the drained state, and a free-energy calculation based on $H(\mu)$ concludes that the robot can perform approximately half a charge of work. This conclusion is incorrect under either outcome of the coin flip: the robot can perform either a full charge of work or none at all. The mean $H(\mu)$ is an average over an ensemble whose members behave nothing like the average, the distribution being bimodal rather than concentrated. The

quantity we actually require is a property of the *individual* state that the robot in fact occupies, and the ensemble formalism defines no such quantity.

Failure of Typicality: Structure Unresolved by the Ensemble. Suppose that the actual state x, while formally a possible sample from the equilibrium ensemble μ, happens to be special in a way that the macrovariables do not capture: the gas particles may, for instance, momentarily occupy a strongly patterned, compressible configuration. The codelength $\hat{H}(x, \mu)$ evaluates x as if it were generic, and a work calculation based on this codelength will *underestimate* what a sufficiently sophisticated machine—one that recognizes and exploits the pattern, that is, a machine running a compression algorithm—can extract from x. Interpreted literally, the ensemble calculation would classify such a machine as a violator of the second law. The correct response is not to prohibit sophisticated machines but to acknowledge that the entropy relevant to what can be done with x is a property of the actual structure of x, not of the ensemble in which we happened to embed it.

6.4 The demon's capacity as the central case for agent foundations

We now return to the demon's free memory, where the difficulty assumes its sharpest form. The demon's remaining optimization capacity is m minus the entropy of its memory contents, and the question immediately arises: *Relative to which distribution is this entropy to be evaluated?* If we, as outside observers, refine our beliefs about what the demon has recorded, the standard formalism asserts that the entropy of the demon's memory drops and that the demon's "capacity to do work" rises, although nothing about the demon's hardware has changed. The classical macrovariable repertoire offers no assistance: temperature, pressure, and chemical composition are designed for systems whose unresolved degrees of freedom mix rapidly and retain no persistent record, whereas a memory is, by engineering design, a system whose states are stable, distinguishable, and meaningful. Every distinct memory configuration must be treated as its own condition; there is no thermalizing ensemble over memory states for μ to attach to, and there is no privileged ensemble over the outputs of a long computation, since the purpose of computation is precisely to produce states whose structure no simple prior anticipates.

For agent foundations, these failure modes are not pathological exceptions to be excluded from the analysis; rather, they constitute the entire subject matter, since the states of primary interest—an agent's memory and world model, the records left by measurements, computations in progress, an environment partway through being reshaped toward a target—are exactly the far-from-equilibrium, information-bearing, intricately structured states on which the ensemble formalism fails. If entropy is to serve as the measure of optimization capacity for embedded agents, we require a definition of entropy that applies to an *individual physical state*, with no exogenous distribution appearing anywhere in the formalism. Such a definition exists, originating in the theory of computation, and its development is the subject of the following section.

7 Algorithmic thermodynamics

Algorithmic thermodynamics, developed by Ebtekar and Hutter on foundations laid by Bennett, Zurek, Gács, and Levin, retains the entire coarse-grained Markov framework of Section 4 while introducing a single fundamental modification: the subjective ensemble μ is replaced by universal computation. Rather than asking how surprising a state x is under μ, the algorithmic framework asks how difficult x is to describe in absolute terms. The conceptual gain is an objective, observer-free entropy that *dissolves the subjectivity problem* of Section 6, being defined for a single physical state without reference to any ensemble. In developing this framework we obtain a second law and fluctuation bounds that hold for individual trajectories, culminating in an exact analysis of Maxwell's demon.

7.1 Kolmogorov complexity

We begin by fixing a universal computer U, which may concretely be regarded as an interpreter for a general-purpose programming language. Every coarse-grained state x of a physical system is identified, via the coarse-graining's encoding, with a finite binary string, so that the notion of a program that outputs x is well defined.

Definition 7.1 (Description Complexity)**.** The *Kolmogorov complexity* (or description complexity) $K(x)$ of a string x is the length, in bits, of the shortest program that outputs x when run on U. The *conditional complexity* $K(x \mid y)$ is the length of the shortest program that outputs x when given y as auxiliary input. The *algorithmic mutual information* between two strings is

$$I(x : y) \;:=\; K(x) + K(y) - K(x, y),$$

the number of bits that a description of one saves in describing the other.[a]

[a] We record two technical refinements once and thereafter suppress them. First, programs are required to be *self-delimiting*: no valid program is a prefix of another, so that each program announces its own end. This convention makes program lengths behave like codeword lengths; in particular they satisfy the Kraft inequality $\sum_x 2^{-K(x)} \leq 1$, which is precisely the property permitting $2^{-K(x)}$ to serve as a (sub)probability distribution. Second, the chain rule for K holds in the form $K(x, y) \stackrel{+}{=} K(y) + K(x \mid y, K(y))$, with the complexity of y itself appearing in the conditioning; consequently some identities below, including the second expression for $I(x : y)$, implicitly carry such terms. Every statement we make is correct with these refinements installed, and none of these refinements is significant at physical scales.

The appropriate interpretation of $K(x)$ is *optimal lossless compression*: the size of the smallest self-contained representation from which x can be regenerated exactly. Several examples serve to calibrate the scale of this quantity. A string of one million zeros has negligible complexity, since a fifteen-byte program suffices to print it. The first million digits of π likewise have negligible complexity, despite passing every statistical test for randomness, because a short program computes them. A string of one million bits obtained directly from a quantum random-number generator has complexity close

to one million bits: almost surely there is no structure to exploit, and the shortest "program" is essentially the string quoted verbatim. A state of 10^{23} particles confined to one corner of a box has low complexity, since a short loop prints the repeated coordinates, whereas a generic thermalized state has complexity near the full length of its encoding. These examples support a crisp characterization of the algorithmic notion: low entropy corresponds precisely to compressibility, a property of the individual state x that makes no reference to any ensemble.

Algorithmic mutual information behaves analogously to the mutual information of Section 2, but for individual objects: $I(x : x) \stackrel{+}{=} K(x)$, since a copy of x renders any further description redundant; $I(x : y) \stackrel{+}{=} 0$ for typical independent random strings, for which a description of one offers no assistance in describing the other; and intermediate values quantify partial correlation, such as that between a measurement record and the system it measured.

Because the shortest program may be arbitrarily difficult to find, statements about K hold up to additive constants reflecting fixed wrapper code; following Ebtekar and Hutter, we write $f \stackrel{+}{<} g$ for $f \leq g + c$, $f \stackrel{+}{>} g$ for $f \geq g - c$, and $f \stackrel{+}{=} g$ for both, where c depends only on fixed choices such as the universal computer, never on the strings involved.

7.2 Justification of K as an entropy measure

We now present three families of results that together justify treating $K(x)$ as *the* entropy of the individual state x.

Physical Negligibility of the Choice of Computer. The quantity K depends on the universal computer U, a dependence that may appear problematic for a proposed physical quantity. The dependence is, however, bounded by translation: for any two universal computers, a fixed interpreter converts programs for one into programs for the other, so the two versions of K differ by at most the interpreter's length, uniformly in x. To gauge the scale of this dependence, we convert to physical units: entropy in bits converts to thermodynamic entropy at 1 bit $= k_{\mathrm{B}} \ln 2 \approx 9.57 \times 10^{-24}$ joules per kelvin. Suppose that two programming languages were so dissimilar that translation between them required a 12-gigabyte interpreter, far larger than any real interpreter. Twelve gigabytes is about 10^{11} bits, so the two languages would assign every physical system entropies agreeing to within $10^{-12}\,\mathrm{J/K}$, far below experimental resolution for macroscopic systems. For all physical purposes, therefore, K may be regarded as computer-independent.

Agreement with the Shannon Framework in Its Domain of Validity. Let μ be any computable probability distribution. Then for *every* state x,

$$K(x \mid \mu) \stackrel{+}{<} \log \frac{1}{\mu(x)} = \hat{H}(x, \mu),$$

because one way to describe x, given access to μ, is to transmit its Shannon codeword (Section 2.3), and the shortest description can only be shorter. In the reverse direction, a counting argument based on the Kraft inequality shows that the inequality is nearly tight for nearly all samples: for any $\delta > 0$, with probability at least $1 - \delta$ over $X \sim \mu$,

$$K(X \mid \mu) \stackrel{+}{>} \log \frac{1}{\mu(X)} - \log \frac{1}{\delta}.$$

In other words, the optimal universal description of a typical sample is its Shannon codeword up to logarithmically small slack; atypical samples may improve upon the Shannon codeword, but only a δ-fraction can do so by more than $\log \frac{1}{\delta}$ bits. Averaging the two displays recovers Zurek's identity

$$H(\mu) \stackrel{+}{=} \langle K(X \mid \mu) \rangle_{X \sim \mu} :$$

the Gibbs–Shannon entropy is the mean Kolmogorov complexity of a sample, given prior knowledge of the distribution. Two specializations sharpen this correspondence: taking μ uniform on a finite set B shows that for any simply describable set B containing x, we have $K(x) \stackrel{+}{<} K(B) + \log |B|$, with near-equality for the vast majority of elements; taking B to be a Boltzmann macrostate (the set of all microstates sharing given macrovariable values) recovers the Boltzmann entropy $\log |B|$ as an upper bound on K, achieved by typical members. Furthermore, the bound applies to *any* simple set rather than to classical macrostates alone: Ebtekar and Hutter's example is that the entropy of a bookshelf can be estimated by taking B to be the set of configurations compatible with the manner in which the books are sorted. The algorithmic entropy thus automatically considers every simple description of the state, classical macrovariables included, and charges the state for the least expensive among them.

These results resolve the three failure modes of Section 6.3. Where the ensemble picture is valid (a simple μ, concentration, and a typical x), we have $K(x) \approx \hat{H}(x, \mu) \approx H(\mu)$, so that algorithmic entropy reproduces the classical answers and nothing is lost. Where the ensemble picture fails, K continues to function correctly. A point mass μ on an intricate state no longer yields zero entropy, because $K(x)$ is large regardless of which distribution is mentioned alongside it, closing the "free knowledge" loophole (the knowledge is now assigned an explicit price, measured in description length). The robot's battery receives a definite per-state answer, K(charged state) or K(drained state), with the ensemble value revealed by Zurek's identity as a mere average of the two. Finally, a secretly patterned gas configuration is credited for its structure, $K(x) \ll \hat{H}(x, \mu)$, so that the sophisticated compression machine extracts exactly the work that the actual description length of x permits, in full accordance with the second law.

The Uncomputability of K and Its Necessity for a Second Law. A fundamental property of K is that no algorithm can compute $K(x)$ from x, a fact whose consequences for the second law we now examine. One direction of approximation remains available: by running ever more candidate programs for ever longer, one obtains a decreasing sequence of upper bounds converging to $K(x)$ (in the standard terminology, K is upper semicomputable), which is why real-world compressors provide genuine

upper bounds on entropy. No algorithm, however, can certify large *lower* bounds on complexity: there is no procedure that, given x, verifiably reports "$K(x)$ is at least one million". This is Chaitin's incompleteness theorem, and its proof is a short self-referential argument: if such a certifier existed, then a short program could enumerate strings until it located one certified to have complexity at least a million and print it, thereby describing a string of supposed complexity at least one million bits with a few hundred bits—a contradiction.

At first sight, uncomputability appears to be a defect in a proposed physical quantity; it is in fact essential to the consistency of the framework. Suppose that thermodynamics were instead formulated with some entropy-like state function K' for which an algorithm *could* certify large values. A short program p could then search for and output a state x^* certified to satisfy $K'(x^*) \gg 0$. However, any short program that can *construct* x^* can also *erase* it: one runs p to produce a second copy of x^* alongside the existing one, uses the copy to reversibly cancel the original (a controlled-NOT for every bit, exactly as in Example 5.1), and then runs p in reverse to restore the auxiliary state. The net effect of a constant-size machine is to take the world from a state containing x^*, with K'-entropy $\gg 0$, to a state not containing it, with K'-entropy ≈ 0. A small, simple, reversible machine that can destroy unbounded amounts of "entropy" at will constitutes a perpetual motion machine with respect to K', and consequently no second law can hold for any computable notion of entropy. The uncomputability of K is precisely what closes this loophole: it guarantees that no simple machine can systematically recognize which states are simple but disguised, and therefore that none can systematically remove the disguise. We note the parallel with Appendix B.8: knowledge that cannot be obtained by any procedure cannot subsequently be expended as a thermodynamic resource. In turn, this justification of K as an objective entropy serves as the foundation for the algorithmic second law, to which we now turn.

7.3 The algorithmic second law

We now return to the physical setting of Section 4: a coarse-grained state space with equal-volume cells, evolving as a Markov chain with doubly stochastic, *computable* transition probabilities (the assumption that the laws of physics admit a short program constitutes a complexity-theoretic version of the Church–Turing thesis, and we choose our reference computer so that the dynamics are simply describable). In this setting the algorithmic entropy of the system in state x is simply $K(x)$, and it obeys a second law descending from Levin's law of *randomness conservation*.

Theorem 7.2 (Algorithmic Second Law; Levin, Ebtekar–Hutter, Stated Informally)**.** *Let (X_t) be a computable, doubly stochastic Markov chain modeling an isolated system's coarse-grained evolution, and let $s < t$ be two times. Then for every $\delta > 0$, with probability greater than $1 - \delta$,*

$$K(X_s) - K(X_t) \overset{+}{<} K(t-s) + \log \tfrac{1}{\delta}.$$

The theorem asserts that *the description complexity of an isolated system's state almost never decreases, except by a precisely priced fluctuation allowance.* It improves upon the ensemble second law (Theorem 4.1) in two respects that are central to our purposes: it holds for each individual trajectory with high probability, rather than merely for the average of an ensemble, and it requires no initial distribution μ whatsoever, removing the exogenous ensemble exactly as Section 6 required.

Each of the two terms in the allowance admits a precise interpretation, and each is quantitatively negligible in physical terms.

The term $\log \frac{1}{\delta}$ prices *chance* fluctuations: rare trajectories along which entropy temporarily falls. Its scale is set by the rarity of the fluctuations one chooses to consider. If events of probability $\delta = 2^{-1000}$ are tolerated, the permitted decrease in entropy is approximately one thousand bits, which in physical units (we recall that 1 bit $\approx 10^{-23}\,\mathrm{J/K}$) is around $10^{-20}\,\mathrm{J/K}$, twenty orders of magnitude below anything measurable. The statistical character of the second law is therefore genuine but quantitatively trivial at macroscopic scales.

The term $K(t-s)$, the complexity of the *elapsed time*, prices *deterministic recurrences*, and its necessity is subtle. The Poincaré recurrence theorem guarantees that an isolated finite system eventually returns arbitrarily close to its initial low-entropy state, so entropy cannot literally always increase. The theorem survives because recurrence times are astronomically long and arithmetically structureless: a system whose recurrence occurs at step $t - s = 17{,}382, \ldots$ (a number comprising some 10^{20} digits) revisits low entropy only at moments whose timestamps are themselves about as complex as the state, and the $K(t-s)$ allowance covers exactly this cost. For any time interval that can be *simply specified* (10^9 years, 2^{100} Planck times), $K(t-s)$ is at most a few hundred bits, so that entropy, regarded as a function of simply specified times, increases monotonically up to negligible slack.

One special case recurs throughout the demon analysis below, and we record it separately.

Corollary 7.3 (Invariance of Entropy under Deterministic Dynamics)**.** *If the evolution over the interval is a computable bijection f with a short description (for instance, a few steps of simply describable reversible dynamics), then*

$$K(f(x)) \stackrel{+}{=} K(x).$$

The reason is immediate from the compression picture: given the dynamics, a description of x is a description of $f(x)$ and vice versa, so the two complexities differ by at most the (small) complexity of f itself. The corollary carries an important consequence: *substantial entropy production requires randomness.* A deterministic, simply describable evolution can shuffle states but cannot increase their complexity by more than a constant; in classical physics, the randomness that drives entropy production is supplied by the coarse-graining of chaotic microdynamics (Section 4.3), which continually injects effectively fresh random bits into the coarse description. A useful picture is the following: a doubly stochastic evolution is equivalent to drawing a random bijection

at each step (for instance, by tossing coins), and entropy can increase by at most the number of coins tossed, and only to the extent that the coins are independent of the state. The Shannon-level second law is blind to this distinction between deterministic reshuffling and genuine randomization, whereas the algorithmic version registers it precisely.

7.4 An exact analysis of Maxwell's demon

Having established the algorithmic second law and its deterministic special case, we now revisit the demon, this time in the form of a theorem rather than of an informal narrative, and recover the Touchette–Lloyd bound as a statement about individual physical states. We take the demon's memory and the gas to be jointly coarse-grained, writing joint states as pairs (memory contents, gas state). The demon's memory begins in a blank reference state 0, and the gas begins in some state x.

Complete Measurement Followed by Erasure. In the idealized case, the demon performs a complete measurement, reversibly copying the gas's state into memory:

$$(0, x) \;\mapsto\; (x, x),$$

and then, using its copy as the control of a conditional operation (a string of controlled-NOTs, Example 5.1), reversibly resets the gas to a reference state:

$$(x, x) \;\mapsto\; (x, 0).$$

Each map is injective on the states that can actually occur (the first on pairs with blank memory, the second on pairs whose two slots agree), hence extendable to a bijection of the joint state space, and each is simply describable. Corollary 7.3 therefore applies to both steps, and the *total* complexity remains unchanged throughout:

$$K(0, x) \stackrel{+}{=} K(x, x) \stackrel{+}{=} K(x, 0) \stackrel{+}{=} K(x).$$

We now examine what this chain of equalities asserts. The gas has passed from complexity $K(x)$ to complexity $\stackrel{+}{=} 0$, which constitutes a complete Type-2 optimization. The joint complexity never changed, so the second law was never threatened at any step, without any need to "complete a cycle" or to invoke ad hoc accounting. The balance is now carried entirely by the demon's memory, which holds a record of complexity $K(x)$; the gas's erasure was permitted *precisely because a copy existed*, since the operation "clear slot two given that slot one holds a copy" is injective. The step that the second law forbids is the clearing of the *last* copy:

$$(x, 0) \;\mapsto\; (0, 0)$$

is either not injective (if it is to work for every x) or not simply describable (if hard-wired for one particular complex x, the wiring itself would have complexity $K(x)$, and Corollary 7.3 charges for the complexity of the dynamics). To recover its memory, the

demon must clear that last copy, irreversibly destroying $K(x)$ bits of complexity; it cannot do so without cost, and the only route consistent with the second law is to export those $K(x)$ bits into the environment as Type-1 waste (Section 5.2). The classic resolution of the demon paradox—that the demon's own information processing is what rescues the second law—emerges here not as an additional postulate but as a direct consequence of the complexity bookkeeping.

Partial Measurement and the Algorithmic Touchette–Lloyd Bound. A realistic demon measures only a small part of the gas state. Let the measurement be $m(x)$, where m is any simply describable (possibly random, possibly many-to-one) function of the gas state—a few bits concerning a single approaching molecule, for instance. The demon records the measurement, then employs the record as a control to drive the gas from x to some new state y:

$$(0, x) \;\mapsto\; (m(x), x) \;\mapsto\; (m(x), y).$$

Whatever feedback protocol the second step implements, provided that it is a simply describable mixture of reversible operations, the algorithmic second law applied to the *joint* system gives

$$K\big(m(x), x\big) \stackrel{+}{<} K\big(m(x), y\big).$$

Expanding both sides with the chain rule ($K(m, \cdot) \stackrel{+}{=} K(m) + K(\cdot \mid m)$, with the technical refinements recorded in Definition 7.1), subtracting $K(m(x))$ from both sides, and rewriting conditional complexities via mutual information ($K(x \mid m) \stackrel{+}{=} K(x) - I(m : x)$), we obtain

$$K(x) - I\big(m(x) : x\big) \stackrel{+}{<} K(y) - I\big(m(x) : y\big) \;\leq\; K(y),$$

where the final step uses the nonnegativity of algorithmic mutual information. Rearranging yields

$$\boxed{K(x) - K(y) \stackrel{+}{<} I\big(m(x) : x\big).} \tag{1}$$

That is, *the gas can lose at most as much entropy as was measured from it.* The budget is, moreover, exactly achievable: if the demon expends its correlation completely and reversibly (so that $K(m(x), x) \stackrel{+}{=} K(m(x), y)$, with no entropy produced, and $I(m(x) : y) \stackrel{+}{=} 0$, with no correlation left unspent), then the displayed inequalities collapse to

$$K(x) - K(y) \stackrel{+}{=} I\big(m(x) : x\big).$$

We now place (1) beside Theorem B.2 (Appendix B) and compare the two statements term by term. In each, the left-hand side is the entropy reduction of the steered system, and the right-hand side is the mutual information between the controller's record (action, measurement) and the system's state. The Touchette–Lloyd theorem is the ensemble (Shannon) version of the constraint, speaking of distributions, averages, and a blind baseline supplied by the dynamics. Inequality (1) is the single-shot,

individual-state (algorithmic) version, speaking of the single gas state actually confronting the demon, without reference to any ensemble, the role of the blind baseline being played by the additive constant, which absorbs what the simply describable dynamics can accomplish unaided. It is simultaneously the rigorous form of Type-3 optimization from Section 5: the copy step purchases $I(m(x) : x)$ and the control step expends it, bit for bit.

One final reading of the demon's situation will prove central to the concluding section. After the measurement, two distinct entropies attach to the same gas: its *objective* entropy $K(x)$, and its entropy *from the demon's perspective*, the conditional complexity

$$K\big(x \mid m(x)\big) \stackrel{+}{=} K(x) - I\big(m(x) : x\big),$$

which is smaller by exactly the measured information. The demon's optimization power over the gas is precisely the wedge between the objective and the subjective entropy. The subjective element has therefore not been eliminated from thermodynamics but rather assigned a physical location: the "subjective distribution" of the older formalism has become a physical conditioning on a physical record, carried in a memory that occupies physical space and obeys the second law.

8 Knowledge as a physical resource: optimization for embedded agents

Having developed the algorithmic account of entropy and examined the channels through which entropy can be displaced, we now synthesize the preceding sections into a single unified picture: for an embedded agent, knowledge about the environment constitutes a physical resource, and this resource is the same quantity as the agent's capacity to optimize the environment.

8.1 From exogenous to endogenous knowledge

We begin by comparing how the two frameworks account for an agent's knowledge of a system.

In the Gibbs–Shannon framework, what an agent knows about a system is encoded in the distribution μ, and μ is supplied from outside the physics, so that entropy, work capacity, and therefore optimization capacity are all defined relative to it. Two problems follow from this exogenous treatment: first, if we change μ, the physics appears to change (the problem of Section 6.4); second, μ is entirely unconstrained, so knowledge can be introduced at no physical cost (the loophole of Section 6.3). At equilibrium neither problem arises, because there the appropriate μ is both forced and simple; the problems become acute precisely when the agent's knowledge is the component of interest.

For an embedded agent no such outside ledger exists. Its knowledge of the environment must be stored physically, in a memory composed of the same matter as the environment and governed by the same laws, and algorithmic thermodynamics enables

us to handle this knowledge directly. Writing a for the physical state of the agent's memory and s for the physical state of the environment, "what the agent knows about the environment" is the algorithmic mutual information

$$I(a : s) \stackrel{+}{=} K(s) - K(s \mid a),$$

the number of bits by which the agent's memory shortens the description of the environment. No distribution appears in this expression: knowledge has become an objective relation between two pieces of matter, as much a fact about the world as a distance or a voltage. Subjective beliefs are not discarded but rather *implemented*: an agent whose memory holds a good model of the environment is one whose state has high $I(a : s)$, and $K(s \mid a)$ is the environment's entropy *as that agent finds it* (the environment's entropy as evaluated from the demon's perspective, Section 7.4).

The Correspondence Between Knowledge and Optimization Capacity. A fundamental insight emerging from this framework is that an agent which understands its environment more thoroughly can effect correspondingly greater change upon it, and the framework renders this intuition precise by joining two bridges: the first passing from the agent's beliefs to mutual information, and the second from mutual information to optimization.

From beliefs to mutual information. Suppose that the agent's memory a encodes a computable belief q about the environment (formally, a short fixed program reconstructs q from a). Given a, one way to describe the true state s is to recover q and record the Shannon codeword of s under q, which has length approximately $\log \frac{1}{q(s)}$ bits (Section 2.3). Since the shortest description can be no longer than this particular one,

$$K(s \mid a) \stackrel{+}{<} \log \frac{1}{q(s)}, \qquad \text{and hence}$$
$$I(a : s) \stackrel{+}{=} K(s) - K(s \mid a) \stackrel{+}{>} K(s) - \log \frac{1}{q(s)}.$$

The more probability the belief assigns to the true environment, the shorter this codeword becomes, and the more mutual information the memory carries about the world.

From mutual information to optimization. The algorithmic Touchette–Lloyd bound (1) establishes that the agent can remove at most $I(a : s)$ bits of complexity from the environment, and that this amount is attainable. Combining the two bridges yields

$$\text{optimization the agent can perform} \stackrel{+}{>} K(s) - \log \frac{1}{q(s)},$$

so the agent can drive the environment down to a residual complexity of approximately $\log \frac{1}{q(s)}$, which is precisely the environment's entropy as the agent's belief regards it. It follows that *the more probability the agent's model assigns to the truth, the more of the environment it can optimize*, at a rate fixed by the bookkeeping rather than by any modeling choice.

This correspondence constitutes Bayesian learning expressed in physical terms. As the agent collects evidence and updates its belief, $q(s)$ rises, the codeword length $\log \frac{1}{q(s)}$ and the residual $K(s \mid a)$ shrink, $I(a : s)$ grows, and the optimization within reach grows with it. In the limit of certainty $K(s \mid a) \stackrel{+}{=} 0$ and $I(a : s) \stackrel{+}{=} K(s)$, and the agent can in principle steer the environment to vanishing residual complexity. Updating a belief toward the truth and accumulating spendable, physically stored knowledge are therefore one and the same process, constituting the single-state form of the Touchette–Lloyd result of Appendix B and the precise content of the Type-3 channel of Section 5.

Two further properties complete the characterization of $I(a : s)$ as a resource.

- It is *consumed* when spent: optimization in this setting is the Type-3 channel (Section 5), in which the shared correlation is consumed, so that continued optimization requires the agent either to measure again (Type 2) or to export waste (Type 1).
- It *decays*: correlations between two systems cannot grow without interaction (a consequence of the algorithmic second law, Section 7.3), so that if the environment changes while the agent's records remain fixed, $I(a : s)$ falls, and correlation that has been allowed to decay can no longer be spent.

This analysis establishes that an embedded agent's knowledge about the world is the same physical quantity as its capacity to optimize the world beyond blind baselines: an agent that knows more can accomplish more, at a fixed exchange rate of one bit of optimization per bit of correlation. In turn, this identification of knowledge with optimization capacity invites a refinement of each of the three channels, to which we now turn.

8.2 Refinements of the three channels via universal computation

Having introduced algorithmic entropy, we now demonstrate that each of the three channels admits a refinement, and in every case the refinement turns on the compressibility of an individual state—a quantity that the ensemble account cannot express.

Type 1. Before exporting waste into the environment, an agent should determine whether the waste is genuinely as random as it appears. If the waste possesses compressible structure (and the output of any computation does, by Corollary 7.3), the agent can compress it first and export only $K(\text{waste})$ bits of genuine entropy rather than its raw size. This compression must occur *before* the waste mixes into the environment, because mixing destroys the structure: once the structure is lost, no compressor can recover it.

Type 2. The memory cost of recording a measured state x is not the raw length of the transcript but its complexity $K(x)$: records can be stored in compressed form, so a memory can absorb more optimization than its nominal size suggests. Moreover, $K(x)$ constitutes a lower bound, in that no method stores x in fewer bits.

Type 3. The steering budget is the mutual information (1), and computation offers a second route to acquiring it: if the environment's state is simple ($K(s)$ small), a short program generates it, so the agent can come to know it (acquiring $I(a : s) \approx K(s)$) by *computation alone*, with almost no measurement. Simple worlds are therefore informationally inexpensive to characterize and consequently inexpensive to optimize.

These three refinements converge in a single fact, noted by Daniel C and Ebtekar: the complexity $K(s)$ of the optimized subsystem is simultaneously the smallest memory that can absorb it (Type 2) and the smallest knowledge that can steer it (Type 3). The cost of recording a state and the knowledge required to steer it are the same number of bits, because both quantities equal the length of its shortest description.

8.3 Dissolution of the apparent subjectivity

We now return to the puzzle of Section 6.4: does the demon's capacity to perform work change when *our* beliefs about its memory change? The answer comprises two parts, of which the second provides the resolution.

First, the demon's capacity is objective: its remaining capacity to optimize is its free memory, namely the size of the memory minus the *description complexity* of its current contents. If those contents are compressible, the demon can compress them (a reversible and costless operation by Corollary 7.3) and thereby recover the space; if they are incompressible, the space is irreducibly occupied, because freeing it would require destroying information, which reversibility forbids. In either case the capacity depends only on the physical state of the memory, and no observer's beliefs enter the formula.

Second, the appearance of belief-dependence nevertheless reflected a genuine physical relation. When we update our beliefs about the demon's memory, that update constitutes a physical change in our own brains, which thereafter share more algorithmic mutual information with the demon's memory. This shared information is itself a Type-3 resource: in principle we could spend it to assist in compressing the demon's memory, freeing capacity that the demon alone could not reach. The demon's situation therefore genuinely differs after our update, but only relative to the enlarged system that now includes our correlated brains, and the difference is exactly the correlation we have gained. What the old formalism recorded as a mysterious observer-dependence of entropy, the new formalism records as an ordinary physical relation between observer and observed, dissolving the apparent subjectivity into objective correlation.

8.4 Thermodynamic implications for optimizing systems

Finally, we reconsider the optimizing systems of Section 3 in the light of these results. An optimizer drives the system upon which it acts from a broad range of configurations into a narrow target set, and the question arises of what the existence of such an optimizer implies thermodynamically.

The funneling constitutes coarse-grained entropy reduction in a subsystem, so the global accounting must remain consistent through some mixture of the three channels (Section 5): waste exported to the surroundings (Type 1), records kept in memory

(Type 2), or pre-existing correlation spent (Type 3). The principal implication is a selection-theorem-like conclusion: if the funneling exceeds the blind baseline of the environment's own dynamics, then by the Touchette–Lloyd theorem (Theorem B.2, Appendix B) and its algorithmic form (1), the system must carry mutual information with what it steers, which is a necessary (though, by Appendix B.8, not sufficient) ingredient of a world model. At least this much modeling is therefore forced by the bookkeeping rather than attributed by an observer. This conclusion constitutes the thermodynamic counterpart of the observer-independence established in Section 3.5: the presence of a world model, like the presence of an optimizer, is an objective fact about the system, certified by the entropy it removes rather than attributed by an external observer.

9 Summary and conclusions

As developed over the preceding sections, our aim in these notes has been to provide a physical, information-theoretic characterization of optimization and to trace its consequences through thermodynamics; we now summarize the principal conclusions. An optimizer is a physical entity whose presence drives the system it acts upon, from a broad range of initial conditions and despite perturbations, into a narrow target: conditioning on the optimizer renders an observer's uncertainty about the system's final state small, endowing the system with a convergent attractor. The amount of optimization is quantified by the entropy reduction the optimizer produces in that system, namely the number of bits by which it narrows the system's possibilities, and the designation of an entity as an optimizer requires no claim about its internals or goals, only about its effect on the system it steers. Even an observer concerned solely with prediction has an objective, information-theoretic reason to distinguish optimizers: because the system converges to the target, the target constitutes a short, robust description of the state in which the system terminates, forecasting the macroscopic outcome for a very small fraction of the bits that a specification of the initial conditions would require, and without any knowledge of the initial conditions at all. This establishes optimization as an observer-independent feature of the world rather than a stance projected by a particular observer, thereby resolving the concern that intelligence is merely task-relative.

Having characterized optimization, we situated it within the constraints imposed by physics. Microscopic physics is reversible, and the second law of thermodynamics follows from that reversibility once the dynamics are coarse-grained: a doubly stochastic Markov step cannot decrease ensemble entropy (Theorem 4.1). Global entropy reduction is consequently impossible; all optimization is local, and the global accounting must remain consistent, with every local reduction compensated elsewhere. Moreover, we established that thermodynamic laws concern information rather than molecules: they bind whenever a designer cannot observe a system, cannot destroy information, and faces a conservation constraint. In Wentworth's coin world, work extraction is data compression, no work can be extracted from a single maxentropic pool, and the fuel of every engine is negentropy—the gap between a system's entropy and its constrained maximum—confirming that the currency of thermodynamic advantage is informational

rather than material.

Building on this foundation, we quantified the informational cost of steering beyond "blind" dynamics. The Touchette–Lloyd theorem (Appendix B) bounds the entropy reduction achievable by any policy by the blind-dynamics baseline plus the mutual information between action and environment, $\Delta H \leq \Delta H^{\max}_{\text{blind}} + I(X;A)$, so that observed optimization in excess of the baseline certifies the presence of modeling; the bound is a constraint rather than a guarantee, since information can be wasted. Complementing this result, we established that an embedded agent has exactly three channels through which it can reduce the entropy of a subsystem: exporting the entropy into the environment as heat, absorbing it into memory by measurement, or spending pre-existing mutual information with the subsystem. Maxwell's demon operates through the second of these channels, its measure-then-steer operation decomposing into the purchase of correlation followed by its expenditure.

Motivated by the demands of embedded agency, we then examined the foundations of the entropy concept itself. The Gibbs–Shannon entropy is defined relative to an exogenous, subjective ensemble, and this dependence fails precisely on the states of which agents are composed: intricate knowledge is assigned zero cost, non-concentrated ensembles yield averages true of no actual state, and atypical structure receives no credit. Equilibrium thermodynamics is the special regime, characterized by simple, concentrated, typical ensembles, in which this failure never manifests.

To address these limitations, algorithmic thermodynamics replaces the ensemble with Kolmogorov complexity, so that entropy becomes an objective property of the individual state, agreeing with the classical entropies within their domain of applicability and extending beyond it. The algorithmic second law holds per trajectory with exactly priced fluctuation allowances (the allowance $K(t-s) + \log \frac{1}{\delta}$ covers Poincaré recurrence and chance decreases in entropy); deterministic simple dynamics cannot create entropy; and the demon obeys $K(x) - K(y) \overset{+}{<} I(m(x) : x)$, the single-state form of the Touchette–Lloyd bound. This objective entropy dissolves the subjectivity problem, being a property of the state rather than of any observer, and the uncomputability of entropy is essential to the framework: a computable entropy admits a perpetual-motion machine.

Finally, for embedded agents knowledge is endogenous: an agent's probabilistic model of its environment is physically encoded in memory, its content is the algorithmic mutual information between the agent's memory and the world, and that mutual information is the consumable, perishable resource that funds optimization beyond blind baselines. Agents possessing greater knowledge of the world are accordingly capable of exerting greater influence upon it, at a fixed exchange rate of one bit of optimization per bit of correlation. Through these results, we have established an account of optimization that is grounded in physics and stated in the language of information, providing both an observer-independent characterization of what optimizers are and a quantitative accounting of what their operation must cost.

A A thermodynamics of biased coins: the generalized heat engine

This appendix develops, as a self-contained worked example, the toy thermodynamic world referenced from two points elsewhere in these notes (the blind baseline of Appendix B and the Type-1 channel of Section 5); it may be read at any point after Section 4. The construction follows Wentworth's essay *Generalized Heat Engine*, whose objective is to take the distinctively thermodynamic ideas of statistical mechanics (heat, work, engines, the impossibility of perpetual motion) and reconstruct them systematically in a setting containing no physics whatsoever, thereby isolating precisely which components of thermodynamics are fundamentally facts about information. The model also makes concrete a perspective, the *designer's viewpoint*, that transfers directly to the analysis of agents embedded in environments they cannot fully observe.

A.1 The designer's viewpoint

We begin by examining the epistemic situation of a designer who wishes to construct a refrigerator, and in particular the question of why the designer cannot simply build a machine that observes the air molecules and removes the fast ones. All of the microscopic dynamics of the combined refrigerator-and-environment system are reversible, so, as established in Section 4.1, the number of possible microstates compatible with what is known never decreases of its own accord, and the only mechanism for reducing uncertainty about a microstate is observation. The designer, however, operates *in advance*: the behavior of the machine is fixed at design time, with no access to the exact positions the air molecules will occupy when the machine is eventually switched on. From the designer's perspective, therefore, there is uncertainty that cannot be reduced but only *relocated*. The machine itself can "observe" variables while running, but the machine is part of the total system, and its observations constitute further reversible dynamics: any record the machine produces is a physical change within the machine, subject to the same conservation rules as every other component of the system. (We analyze precisely this maneuver, observation implemented as reversible copying, in our treatment of Maxwell's demon in Section 5.)

The design problem is therefore to choose transformations, in advance and without inspection of the system, that make some chosen variables (the interior of the refrigerator) *more* certain at the cost of making others (the heat baths powering the machine) *less* certain. Thermodynamic-style laws apply whenever three conditions hold: the designer cannot gain information about the system (no "peeking" at design time), information cannot be destroyed at the lowest level (reversibility), and some conserved quantity constrains the admissible transformations (energy, in the physical case). Wentworth's toy world possesses exactly these three conditions and nothing else, providing a minimal setting in which the resulting laws can be derived in full.

A.2 Specification of the model: coins, transformations, and conservation laws

Having identified the three conditions under which thermodynamic-style laws apply, we now specify Wentworth's toy world in full. The world consists of two large pools of independent biased coins, in which each coin shows 1 (heads) or 0 (tails):

- a *cold pool* $X_1^C, \ldots, X_n^C$, in which each coin is heads with probability 0.1 (entropy $h(0.1) \approx 0.47$ bits per coin, by Example 2.2); and
- a *hot pool* $X_1^H, \ldots, X_n^H$, in which each coin is heads with probability 0.2 (entropy $h(0.2) \approx 0.72$ bits per coin).

The heads-probability plays the role of temperature (the hot pool is more uncertain per coin), while the *number of heads* plays the role of energy. We may apply any transformation T that replaces the coins' values with new values computed from their old values, subject to three rules mirroring the three conditions identified above:

1. **Reversibility.** T must be invertible: from the final state of all the coins (together with knowledge of which transformation was applied), the initial state must be reconstructible. This is the analogue of microscopic reversibility.
2. **Conservation.** T must conserve the total number of heads, the analogue of energy conservation: heads may be relocated between coins but neither created nor destroyed on net.
3. **Design-Time Ignorance.** The transformation T is chosen before any coin is inspected (the "no peeking" condition), exactly as the refrigerator designer commits to a blueprint before the machine encounters its environment.

To demonstrate that interesting transformations exist within these rules, we consider Wentworth's example, in which T acts on three coins as follows: *if* X_1^C *shows tails, swap* X_3^H *with* X_7^H*; if* X_1^C *shows heads, change nothing.* This transformation conserves heads (either two coins are swapped, which permutes values without changing their sum, or nothing occurs), and it is reversible (the control coin X_1^C is itself unchanged, so the final state determines whether the swap occurred, and applying the same transformation a second time undoes it). Conditional swaps of this kind, composed in sequence, constitute the fundamental building blocks from which all admissible engines are assembled, and the example already establishes a fundamental point: a transformation can *read* one part of the system and act on another, entirely within deterministic, reversible, conservative rules, demonstrating that machines that "observe while running" lie not outside the formalism but within it.

In summary, the designer selects an invertible, heads-conserving map T from coin configurations to coin configurations, the world applies it, and our objective is to characterize precisely what such maps can accomplish.

A.3 Work extraction as a compression problem

We define a *work coin* to be a coin whose final value is heads with near-certainty (probability approaching 1 as n grows). Work coins are the analogue of stored mechanical work or a charged battery: a resource in a known, definite state, available to power other processes. *Extracting work* means choosing T so that some w designated coins are rendered work coins by the transformation, thereby converting statistical uncertainty into a resource in a known state.

The first key observation is that reversibility transforms work extraction into a data compression problem. We write X for the (random) initial configuration of all the coins and $X' = T(X)$ for the final configuration. Because T is invertible, the final state determines the initial state: X' contains *all* of the information in X, so $H(X') \geq H(X)$ (in fact $H(X') = H(X)$, since invertible maps preserve entropy exactly). Suppose now that w of the final coins are deterministic. Deterministic coins carry zero entropy, so all $H(X)$ bits of initial uncertainty must be carried by the remaining coins; in other words, *extracting w work coins requires compressing the information content of the entire initial state into the other coins.* Producing certainty in one location therefore requires concentrating uncertainty elsewhere, reflecting the fundamental fact that under reversible dynamics nothing is ever erased.

A.4 The impossibility of work extraction from a single heat bath

We now attempt to extract work from the hot pool alone and demonstrate that the attempt must fail; this failure constitutes the toy-world analogue of the impossibility of a perpetual motion machine of the second kind (a machine that extracts work from a single-temperature heat bath).

The hot pool's n coins carry $H(X) = 0.72n$ bits of entropy, and a compression scheme can in principle encode these bits into approximately $0.72n$ coins, leaving $0.28n$ coins deterministic. The conservation law, however, renders this compression infeasible, as the following counting argument demonstrates. Fully compressed data is incompressible, which means that it appears statistically uniform: each compressed coin is heads with probability approximately $\frac{1}{2}$ (if the compressed coins were biased, they would admit further compression, contradicting full compression). The $0.72n$ compressed coins would therefore contain approximately $0.36n$ heads, whereas the initial state contains only $0.2n$ heads, and heads are conserved. Even if every one of the intended work coins were set to tails rather than heads, the accounting cannot be made consistent: the required compression demands more heads than the system possesses, and the attempt fails.

This argument admits a generalization extending well beyond the setting of coins, which we now develop in order to characterize the resource underlying all work extraction. The hot pool's distribution (independent coins, each heads with probability 0.2) is the *maximum-entropy* distribution among all distributions with its expected number of heads; in this precise sense a heat bath is "as random as its energy allows". Suppose, in general, that a pool of variables is maxentropic subject to a constraint that fixes the value of some additive quantity $\sum_k f_k(X_k)$. Deterministically fixing the value of

one variable (extracting work from it) shrinks the set of values the constrained sum can take on the remaining variables, and therefore shrinks the maximum entropy the remaining variables can hold. Since the initial state already saturated this maximum, the remaining variables cannot absorb all of the information, and compression must fail. *Work cannot be extracted from a system that is already at maximum entropy given its constraints.* The slack between a system's actual entropy and its constrained maximum, which following standard usage we call *negentropy*, is therefore the sole resource from which work can be drawn.

A.5 Work extraction from two heat baths at different temperatures

We now employ both pools: $2n$ coins, with total entropy $\approx 0.47n + 0.72n = 1.19n$ bits and total heads $\approx (0.1 + 0.2)n = 0.3n$. The crucial difference from the single-pool case is that the *joint* initial distribution is not maxentropic given the joint constraint: the maximum-entropy distribution with $0.3n$ heads spread over $2n$ coins would make every coin heads with probability 0.15, whereas the actual state maintains two distinct biases, 0.1 and 0.2. The gap between actual entropy and constrained maximum entropy is negentropy, and it constitutes a resource that the designer can expend.

We now quantify the amount of work that this negentropy affords, deriving the maximal number of work coins extractable from the two pools. Suppose that we designate w coins as intended work coins (deterministic heads). The remaining $2n - w$ coins must carry all $1.19n$ bits of initial entropy while containing the remaining $0.3n - w$ heads. To carry as much information as possible, the final distribution of those $2n - w$ coins should itself be maxentropic subject to its heads constraint, which (for large n) means independent coins, each heads with probability

$$q \;=\; \frac{0.3n - w}{2n - w},$$

giving total entropy $(2n-w)\,h(q)$ where h is the binary entropy function of Example 2.2. The largest feasible w makes this capacity exactly equal to the required $1.19n$ bits:

$$(2n - w)\; h\!\left(\frac{0.3n - w}{2n - w}\right) \;=\; 1.19n.$$

Solving numerically gives $w \approx 0.011n$. A heads-conserving, reversible, designed-in-advance transformation therefore *can* extract work from two heat baths at different temperatures, converting approximately 3.7% of the total heads (5.5% of the hot pool's heads) into work coins. This construction is the toy world's heat engine, and its derivation required nothing beyond information-theoretic reasoning, confirming the claim that the distinctive behavior of heat engines is fundamentally a fact about information.

Wentworth notes one caveat regarding comparison of this figure with the physics literature: the efficiency computed here is not the classical Carnot efficiency, because the classical result answers a slightly different question (the optimal conversion rate *at the margin*, for engines drawing from effectively inexhaustible baths, generally consuming unequal amounts from the hot and cold sides), whereas we have asked how much total

work can be extracted from two fixed, finite pools. The conceptual content—that work arises only from temperature *differences* and never from a single bath—is identical.

A.6 Lessons from the toy world

Four lessons from this construction recur throughout the remainder of these notes, and we therefore state them explicitly.

Substrate Independence of Thermodynamic Law. Thermodynamic laws are not specifically about molecules and joules; they apply whenever an agent or designer cannot gain information about a system at will, cannot destroy information at the lowest level, and operates under some conservation constraint. This is the reason the same laws reappear when we analyze agents, memories, and computations.

Conservation of Uncertainty under Reversibility. Under reversibility, uncertainty is never destroyed but only relocated, so that every machine that makes one variable more predictable necessarily makes others less predictable.

Work Extraction as Compression. Extracting work (producing variables in definite, known states) is a compression problem, and conversely, compression capacity is work capacity. This identity between "useful energy" and "compressibility" is the seed of the algorithmic viewpoint of Section 7, where it is sharpened from a statement about distributions into a statement about individual states.

Negentropy as the Sole Fuel. The fuel of all such machines is negentropy: the gap between a system's entropy and the maximum entropy compatible with its constraints. A system at its constrained maximum entropy (a single heat bath) is useless as fuel, regardless of how much "energy" it contains.

Everything in this section was achieved *blind*: the transformation was chosen before any coin was inspected, and all of the extracted work derived from statistical structure (the bias difference between the pools) that was known at design time. This raises the natural question of what an agent could additionally accomplish if it were permitted to *observe* the system before acting, a question that admits a precise answer and that we take up in Appendix B.

B The informational cost of steering: the Touchette–Lloyd theorem

B.1 From optimization to modeling: the motivating question

Having established in Section 5 that an agent can steer a subsystem by expending mutual information it already shares with that subsystem (the Type-3 channel), we now quantify this channel precisely, asking how much entropy reduction each bit of mutual information actually purchases. The answer establishes a rigorous connection between optimization and *modeling*.

One of the recurring aspirations of agent foundations is the identification of *selection theorems*: results establishing that any system selected to perform sufficiently well at some task must, as a matter of mathematical necessity, contain certain agent-like

structures, such as a world model, a goal representation, or a planning process. The *agent structure problem* poses this question in the converse direction to the usual one: rather than asking whether agents bring about outcomes, it asks whether a system observed to reliably bring about a particular outcome must necessarily be modeling its environment. Wentworth formulated a sharp quantitative version of this question: how many bits of optimization can one bit of observation purchase?

A theorem of Touchette and Lloyd, originally published in the control theory literature and brought to the attention of the agent foundations community by Harwood and Altair, constitutes one of the few existing results that directly addresses this problem.[2] Informally, the theorem states that *the entropy reduction a policy achieves, beyond what the environment's dynamics would accomplish on their own, is bounded by the mutual information between the policy's action and the environment's state*, so that every bit of optimization beyond the blind baseline requires a corresponding bit of modeling of the environment's state. This appendix develops the exact statement systematically, accompanied by fully worked examples.

B.2 The formal setup: environments, actions, and policies

The model comprises a single time step of an environment subject to external influence, involving three random variables:

- X, the *initial state* of the environment;
- A, the *action* taken (by an agent, a controller, or a machine; the formalism is indifferent to the nature of the actor);
- Y, the *final state* of the environment.

Two conditional distributions specify the situation completely. The *policy* $P(A \mid X)$ describes how the action is chosen as a function of the environment's initial state, and the *dynamics* $P(Y \mid X, A)$ describe how the final state is produced from the initial state together with the action. The dynamics are held fixed, representing the physics of the environment; the policy is the object of study. As is conventional, X and Y range over the same set of environment states, and the dynamics may be deterministic (all transition probabilities 0 or 1) or noisy.

Following Section 3.3, we score a policy by the entropy reduction it achieves:

$$\Delta H \;:=\; H(X) - H(Y),$$

[2]The result appears as Theorem 10 of H. Touchette and S. Lloyd, *Information-theoretic approach to the study of control systems*, Physica A 331:140–172 (2004); it is also described in their 2000 paper *Information-theoretic limits of control*, and an earlier, more physics-flavored proof appears in Lloyd's 1989 work on the use of mutual information to decrease entropy, in the context of Maxwell's demon. Related results connecting regulation to modeling include the good regulator theorem of Conant and Ashby and the internal model principle of control theory, but these concern conditions for *optimal* regulation, whereas the Touchette–Lloyd theorem is an inequality constraining *all* policies, optimal or not, which is precisely the property that qualifies it as a genuine selection theorem.

the number of bits by which the final state is more predictable than the initial state. This measures optimization precisely in the manner prescribed by Section 3, namely by the extent to which the process funnels a broad distribution into a narrow one. (As noted in Remark 3.1, entropy reduction constitutes the physics-facing half of expected utility maximization, the other half being the specification of which narrow region the agent prefers.)

B.3 Blind and sighted policies

Definition B.1 (Blind and Sighted Policies)**.** A policy is *blind* if the action is statistically independent of the initial state, that is, if $P(A \mid X) = P(A)$, or equivalently $I(X; A) = 0$. A policy is *sighted* if $I(X; A) > 0$.

Blind policies include every deterministic rule of the form "always take action a_1", and also every randomized rule whose randomness is independent of the environment, such as "flip a private coin; on heads take action a_1, on tails take action a_2". What blindness excludes is precisely any flow of information from the environment's state into the choice of action. Sightedness is a matter of degree, and the degree is measured by $I(X; A)$: a policy that conditions on one observed bit of the state has $I(X; A) \leq 1$, while a policy that observes everything can have $I(X; A)$ as large as $H(X)$. The entire repertoire of transformations available in the coin world (Appendix A) consisted of blind policies: the "no-peeking" rule was exactly the requirement $I(X; A) = 0$, with the "action" being the choice of transformation.

One might naively conjecture that any entropy reduction whatsoever requires sight, but this conjecture is false, and understanding why it fails sharpens the eventual theorem. The dynamics alone can reduce entropy: a contracting dynamics (one that funnels many initial states toward the same point on its own, as a ball settles to the bottom of a valley against friction) funnels states regardless of the action taken, and the coin engine of Appendix A.5 reduced the entropy of designated coins while remaining completely blind. The correct question is therefore not whether a policy reduces entropy, but whether it reduces entropy *beyond what blindness allows*. Accordingly, we define the *blind baseline*

$$\Delta H^{\max}_{\text{blind}} := \max_{P(X)\in\mathcal{X},\, P(A)\in\mathcal{A}} \Delta H,$$

the largest entropy reduction achievable by *any* blind policy from *any* initial distribution, where $\mathcal{X}$ is the set of all distributions over initial states and $\mathcal{A}$ is the set of all action distributions independent of X. This baseline is a property of the dynamics alone, capturing everything the environment can be induced to do "on its own".

B.4 A worked example: the guessing game under blind play

The following game, drawn from Harwood and Altair's exposition, renders every quantity in the theorem concrete and computable.

A computer secretly selects a 5-bit binary string X uniformly at random, so that $H(X) = 5$ bits. The player then submits a 5-bit string A. The computer feeds both strings into the fixed, publicly known function

$$f(x,a) \;=\; \begin{cases} \texttt{00000} & \text{if } a = x \\ & \quad \text{(the submitted string matches the secret string)}, \\ x & \text{otherwise}, \end{cases}$$

and outputs $Y = f(X, A)$. The player's objective is to render the output distribution as predictable as possible, that is, to minimize $H(Y)$.

We first consider blind play, in which the string must be chosen with no knowledge of X, and ask how much entropy reduction can be achieved. Consider the submission of a fixed string a. If the player chooses $a = \texttt{00000}$, then a match produces the output 00000 while a failure to match reproduces X, which (conditional on not matching) is uniform over the 31 strings other than 00000; a short calculation shows that the output is then exactly uniform over all 32 strings, so that $H(Y) = 5$ bits and no reduction whatsoever is achieved. The choice $a = \texttt{00000}$ is therefore the unique submission that yields no entropy reduction, and we exclude it from consideration in what follows.

Suppose instead that the player submits any other fixed string, say $a = \texttt{11111}$. Two initial states lead to the output 00000: the state $X = \texttt{11111}$ (a match, upon which the function outputs zeros) and the state $X = \texttt{00000}$ (no match, upon which the function reproduces it). The output 11111 itself never occurs (if $X = \texttt{11111}$ the player has matched it, producing zeros; otherwise the output is $X \neq \texttt{11111}$). Every other string y occurs exactly when $X = y$. The output distribution is therefore

$$P(Y = \texttt{00000}) = \tfrac{2}{32} = \tfrac{1}{16}, \qquad P(Y = \texttt{11111}) = 0,$$
$$P(Y = y) = \tfrac{1}{32} \;\text{ for the other 30 strings},$$

with entropy

$$H(Y) \;=\; \tfrac{1}{16}\log 16 \;+\; 30 \cdot \tfrac{1}{32}\log 32 \;=\; 0.25 + 4.6875 \;\approx\; 4.94 \text{ bits}.$$

This represents a modest but genuine improvement over 5 bits: two initial states have been funneled onto a single output, rendering the output distribution slightly non-uniform and therefore slightly more predictable. By symmetry, every fixed non-zero string performs identically, randomization among them yields no further improvement, and one can verify that this strategy is optimal among blind policies; for this game and this initial distribution, the blind optimum is $\Delta H \approx 0.06$ bits. (If the initial distribution over strings were non-uniform, the optimal blind strategy would be to submit the most probable string other than 00000; the theorem's baseline $\Delta H^{\max}_{\text{blind}}$ takes the maximum over initial distributions as well.)

B.5 The guessing game under sighted play

We now suppose that, before choosing the submitted string, the player is shown the first k bits of the computer's secret string. A natural strategy is to submit the string

consisting of the k observed bits followed by all `1`s (the continuation is largely immaterial, provided the player avoids completing the all-zeros string). Observing k bits multiplies the probability of an exact match by 2^k, from $\frac{1}{32}$ to $\frac{1}{2^{5-k}}$, and each match funnels probability onto the single output `00000`. Computing $H(Y)$ for each k exactly as above produces the following table.

bits of X observed	0	1	2	3	4	5
$H(Y)$ in bits	4.94	4.85	4.63	4.11	2.83	0
ΔH in bits	0.06	0.15	0.37	0.89	2.17	5

With all five bits observed, the player matches the secret string on every round, the output is deterministically `00000`, and the maximum conceivable reduction $\Delta H = H(X) = 5$ bits is achieved. Each additional bit of observation purchases additional steering power, with the marginal returns increasing in this particular game as more of the state becomes known, demonstrating that information about the environment converts into optimization of the environment.

B.6 Mutual information as a measure of sightedness

To state the theorem, we must quantify the degree to which a policy is sighted, and the appropriate quantity is precisely the mutual information of Definition 2.6. In the strategy above with $k = 2$, the action is determined by the two observed bits (the player always submits the two observed bits followed by `111`). Consider an observer who knows this strategy and observes only the *action*, say $A =$ `10111`. The observer can immediately infer that the secret string begins with `10`: the action reveals the observed bits. Before seeing the action, the observer's uncertainty about X was $H(X) = 5$ bits; after seeing it, $H(X \mid A) = 3$ bits; hence $I(X; A) = H(X) - H(X \mid A) = 2$ bits, exactly the number of bits the policy "knows". Mutual information thus formalizes the intuitive notion of how many bits of the environment's state are reflected in the agent's behavior, and it does so without any reference to the agent's internals: it is a property of the joint statistics of state and action, estimable in principle by observing the system over many rounds of play.

B.7 Statement of the theorem

Theorem B.2 (Touchette–Lloyd)**.** *Fix any dynamics $P(Y \mid X, A)$. Then for every initial distribution $P(X)$ and every policy $P(A \mid X)$,*

$$\Delta H \;\leq\; \Delta H^{\max}_{\text{blind}} \;+\; I(X; A).$$

In words, the entropy reduction achieved by an arbitrary policy decomposes into at most two budgets: everything the dynamics could have been steered to accomplish by a blind controller, plus one bit for every bit of mutual information between the action and the environment's initial state. The theorem holds with no assumptions about

the policy's internal structure, no optimality requirements, and no restriction on the dynamics. In particular, *it makes no reversibility assumption*: it is a purely information-theoretic (data-processing) inequality, valid for arbitrary dynamics, reversible or not. Far from standing in tension with the remainder of the development, this is consistent with it: the second law of Section 4.4 required reversibility (in the form of double stochasticity) to forbid *global* entropy reduction, whereas the Touchette–Lloyd theorem requires no such assumption to bound the entropy reduction of a steered subsystem *beyond the blind baseline*. The two results constrain different quantities, and both hold simultaneously.

The contrapositive direction endows the theorem with its selection-theorem character. Suppose we observe a system achieving an entropy reduction ΔH strictly exceeding the blind baseline of its environment's dynamics. We may then *deduce*, without any inspection of the system's internals, that the system's actions carry at least $\Delta H - \Delta H^{\max}_{\text{blind}}$ bits of mutual information with the environment's state. Mutual information with the environment is plausibly a necessary (though certainly not sufficient) ingredient of anything deserving the name *world model*; the theorem thus constitutes a first rigorous step along the path from observed optimization to internal modeling, which is the path toward understanding when optimization implies agent-like structure. In the vocabulary of the guessing game, any player who achieves an output entropy below 4.94 bits must have observed some portion of the secret string, with the margin of improvement providing a lower bound on the number of bits observed.

B.8 Limitations of the theorem

The inequality provides an upper bound on what information makes possible—not a guarantee that information will be exploited effectively—and this gap manifests in both directions.

First, mutual information may simply fail to confer any advantage. If the dynamics ignore the action entirely (Y depends only on X), then every policy, however well informed, performs exactly as a blind one does: the channel from action to environment has zero capacity, rendering knowledge without influence entirely inert.

Second, mutual information can be actively squandered. In the guessing game, consider the policy that observes the secret string completely and submits its bitwise negation (if the computer selects `01100`, the player submits `10011`). This policy attains the maximal $I(X; A) = 5$ bits of mutual information with the environment, and yet it *never* matches the secret string, so the output is always $Y = X$, uniformly distributed: $\Delta H = 0$, strictly worse than the optimal blind policy. This demonstrates that possession of maximal mutual information with the environment is entirely compatible with arbitrarily poor steering performance, confirming that modeling constitutes a necessary rather than a sufficient condition for optimization.

The theorem should consequently be read as a conservation-style constraint in the same family as the second law: it specifies what cannot happen (substantial optimization without modeling), never what must happen (modeling producing optimization). The analogy is exact in spirit, and Section 7 converts it into a literal theorem of ther-

modynamics, in which the role of $I(X;A)$ is played by the mutual information between a measurement record and the measured system.

Example B.3 (Noise-Canceling Headphones)**.** An everyday system exhibits the complete structure of the theorem. Let X be the ambient sound arriving at a listener's ears, let Y be the sound the listener actually hears, and consider two technologies. Foam earplugs attenuate sound by passive damping. They are entirely blind ($I(X;A) = 0$; the plug's "action" is identical whatever the sound), and yet they achieve a substantial entropy reduction: they exploit fixed statistical structure of the environment (sound is vibration, and foam damps vibration), constituting exactly a blind policy operating within the $\Delta H^{\max}_{\text{blind}}$ budget, in the same manner as the coin engine operating on the known bias difference. Active noise-canceling headphones operate in a categorically different manner: they *listen* to the incoming sound and emit its inverted waveform. The emitted signal carries high mutual information with the ambient sound, and the entropy reduction correspondingly exceeds anything passive damping can achieve; the headphones can even steer selectively, canceling the drone of an engine while transmitting a human voice. Finally, playing music through the headphones constitutes an entropy-*increasing* action, demonstrating that agents are under no obligation to minimize entropy; the theorem merely prices the steering they elect to perform.

Remark (The Coin Engine, Revisited)

The generalized heat engine of Appendix A and the Touchette–Lloyd theorem constitute two halves of a single picture. The engine demonstrates what blind policies can extract: everything within the $\Delta H^{\max}_{\text{blind}}$ budget, which is funded by statistical disequilibrium known in advance (the temperature difference between the pools). The theorem prices what blindness cannot reach: every further bit of entropy reduction costs a bit of mutual information acquired through observation. This provides the quantitative form of the Type-3 channel of Section 5: $I(X;A)$ is precisely the correlation an agent expends when it steers, and Section 7.4 demonstrates, with the accounting balancing exactly, what acquiring this correlation costs in turn.

Sources and further reading

These notes integrate the following sources, listed in approximately the order in which their material appears.

- A. Flint, *The ground of optimization*, AI Alignment Forum, 2020. A complementary descriptive treatment of optimization in terms of *optimizing systems* (a broad basin of attraction, a narrow target set, and robustness to perturbation), together with a finer-grained comparison of such systems along the axes of robustness,

duality, and retargetability. Recommended as further reading on the behavioral characterization of Section 3.

- E. Yudkowsky, *Measuring optimization power*, LessWrong, 2008. A quantitative proposal that measures optimization by the improbability of the achieved outcome under random rearrangement, which may be read alongside the entropy-reduction measure of Section 3.3.

- D. H. Wolpert and W. G. Macready, *No free lunch theorems for optimization*, IEEE Transactions on Evolutionary Computation 1(1):67–82, 1997; and D. C. Dennett, *The Intentional Stance*, MIT Press, 1987. The two articulations of the observer-relative view of intelligence and agency that Section 3.5 presents and addresses: no optimizer is universally competent (competence is relative to a problem class), and agency is a predictive stance adopted by an observer rather than an intrinsic property.

- J. Wentworth, *Generalized heat engine*, LessWrong, 2020. The source for the entirety of Appendix A, including the designer's viewpoint, the biased-coin world, work extraction as compression, and the two-bath engine. His decomposition of expected utility maximization (Remark 3.1) appears in *Utility maximization = description length minimization*, LessWrong, 2021.

- A. Harwood and A. Altair, *When bits of optimization imply bits of modeling: the Touchette–Lloyd theorem*, LessWrong, 2025. The pedagogical source for Appendix B, including the guessing game, the blind and sighted vocabulary, the headphone analogy, and the caveats concerning insufficiency. The underlying theorem is due to H. Touchette and S. Lloyd, *Information-theoretic approach to the study of control systems*, Physica A 331:140–172, 2004 (Theorem 10), with antecedents in their 2000 paper *Information-theoretic limits of control* and in Lloyd's 1989 work on Maxwell's demon.

- Daniel C and A. Ebtekar, *Algorithmic thermodynamics and three types of optimization*, AI Alignment Forum, 2025. The central organizing source for these notes, providing the characterization of optimizers through the convergent attractors they create and the information-theoretic argument for attending to such attractors irrespective of one's goals (Section 3), the three types of optimization (Section 5), the argument that entropy should objectively measure optimization capacity (Section 6), the algorithmic refinements of the three types, and the embedded-agency synthesis (Section 8).

- A. Ebtekar and M. Hutter, *Foundations of algorithmic thermodynamics*, Physical Review E, 2025 (arXiv:2308.06927). The formal backbone of Sections 4 and 7, supplying Markovian coarse-grainings and the multibaker construction, Gács' coarse-grained algorithmic entropy, Levin's conservation of randomness and the algorithmic second law with its $K(t-s)+\log\frac{1}{\delta}$ allowance, the exact analysis of Maxwell's

demon including the partial-measurement bound (1), and the ensemble-versus-state comparison of Section 6 (including the robot-battery and bookshelf examples and Zurek's identity).

For the broader agent foundations context assumed in Section 1 (true names and Goodhart's law, reflective stability, embedded agency, selection theorems), we refer the reader to the Agent Foundations slide deck accompanying this module.