The Publication Decision in the Calculus of Constructions:

Rebound Effects, Gamed Proxies, and a Hypothesis That Remains Unproved

Flyxion
Independent Researcher

October 2026

Abstract

Two recent essays describe a procedure for drawing sources from a large corpus, proposing an argument from them, and developing the argument under recorded obligations. The question of whether to publish them was debated informally in a third essay, which left the matter open. The present essay formalizes the question. It states the hypothesis H that publishing the two essays is warranted, develops a dependent type theory in the style of the calculus of inductive constructions together with a small process algebra in the style of SpherePOP, and proves a sequence of lemmas and theorems about the rebound effect named for Jevons, about proxies that are gamed in the manner described by Goodhart, about the counting of combinations, about the case for analyzing even unlikely risks, about dependence on a provider that hosts the author's identity and repositories, about the meaning of a contribution count, about the independence of the hypothesis from the stated axioms, and about the self-endorsing sentences of Löb. None of these results proves H. None of them proves the negation of H. The essay proves instead that H is equivalent, under its own axioms, to a conjunction of three editorial conditions, each of which is an atomic proposition the axioms do not decide, and it states this fact in every section so that no reader mistakes the formal apparatus for a verdict. The formal development has not been mechanically checked, and every theorem in the accompanying proof-assistant listing is closed with the command `Admitted`. The satire, to the extent that there is any, consists of the apparatus being larger than the conclusion it supports.

1 The Hypothesis and Its Obstacles

Two essays exist in draft. The first, on the recombination of a corpus of several hundred essays into briefs for new ones, counts the unordered pairs and triples of the corpus, specifies a short generator that samples them, and argues that the count measures the corpus's capacity for juxtaposition and not its capacity to yield essays. The second specifies the pipeline around the generator: four independent organizations of the corpus, a staged sequence of role prompts to one local model, a coordinating decision on readiness, and composition accepted by a ledger of obligations and not by a score. Neither essay reports a measured result. The generator's call to the local model has not been exercised in the environment where it was written, and the worked example is an editorial one.

The hypothesis under study is the following.

Definition 1.1 (The hypothesis)**.** H is the proposition that publishing the two essays in the public repository in their present state, with the editorial changes available to the author, is warranted.

The word *warranted* is left informal in Definition 1.1, and Section 13 shows that nothing in the development requires it to be made precise, because the development never establishes H and never refutes it. This is the first of the essay's three obstacles: the hypothesis is not decided by the axioms that the author is prepared to state.

The second obstacle is the rebound effect. Jevons observed in 1865 that improvements in the efficiency with which coal was used tended to increase and not decrease the total consumption of coal, because cheaper use enlarged the set of uses worth pursuing [4]. A generator that makes briefs cheap to produce is an efficiency improvement of exactly this kind, and the resource it makes cheaper is not the only resource it consumes. The brief is cheap to generate and expensive to read, and a policy of generating many briefs raises the burden on the one reader the procedure cannot do without. Publishing the essays advertises the procedure and so tends to increase its use. Whether that increase is a reason for or against publication depends on a parameter that the essay does not know, namely the elasticity of the demand for briefs, and the essay says so.

The third obstacle is the gaming of metrics. Goodhart's remark, made about monetary policy in 1975, that a statistical regularity tends to collapse once pressure is placed on it for control purposes, has been restated in many forms, including Campbell's law about social indicators and Strathern's formulation that a measure which becomes a target ceases to be a good measure [3, 1, 10]. The pipeline described in the two essays replaces numerical scores with a ledger of obligations because the scores proposed in the earlier record were gameable. The ledger is itself a metric, and a model that both drafts a section and proposes that its obligations are discharged is in a position to optimize the count of discharged obligations by the cheapest available route. The essays acknowledge this. The formal development below shows how far the acknowledgment goes and where it stops.

Two further considerations were raised once the first three had been drafted, and both concern the world outside the repository. The first is dependence on a provider: the author's accounts with a large provider of storage and identity services serve as the means of signing in elsewhere, the repositories are hosted on a platform whose alternative the author has contemplated building, and an essay in the repository is critical of the sort of provider on which these dependencies rest. The footprint is small, and the exposure, as Section 11 shows, does not depend on the footprint. The second is the contribution graph, whose meaning depends on a convention about how many commits constitute a normal quantity of work, and which is therefore a metric in the sense of the third obstacle. These considerations require a justification for analyzing events that may be unlikely, and Section 10 gives one that does not depend on an estimate of any probability.

Each obstacle is treated by a theorem. The theorems are proved in the ordinary mathematical prose of the metatheory, and the same statements are given in the notation of a proof

assistant in Appendix A, where, in the interest of candor, every proof is closed with `Admitted`. A development closed in that way has the form of a proof and the status of an assumption. The prose proofs are the actual proofs, and a reader who doubts one should doubt it there.

The convention of disclaimers

Every section ends with a boxed statement of the status of H. The statement is the same in substance each time: the hypothesis has not been proved. It is repeated because the formal register invites a reader to infer that something has been demonstrated, and what has been demonstrated is always something other than H. The repetition is not a stylistic flourish added to a finished argument. It is the finished argument.

> Status of H. Unproved. This section states the hypothesis and the three obstacles. It proves nothing about H.

2 Summary of Results and Non-Results

Table 1 lists the statements the essay proves, the statements it states without proof, and the statements it declines to prove. The third column names the section in which the statement occurs. The final entries are the ones the essay was written to be unable to supply.

The proved items are conditional on assumptions that the essay states where they are used, and the assumptions are collected in Section 13, where they become the axioms of the independence result. A reader who accepts all of them has proved R1 through R19. A reader who accepts all of them has not thereby come any nearer to N1 or N2, and the independence result in Section 13 is the proof that no reader could.

> Status of H. Unproved. The table lists H as item N1 and $\neg H$ as item N2. Both are marked not proved, and the remainder of the essay is a development of why.

3 The Type Theory

The object theory is a calculus of inductive constructions in the tradition that begins with the calculus of constructions of Coquand and Huet [2] and acquires inductive definitions in the system of Paulin-Mohring [9]. The presentation below is deliberately minimal. Universe levels are suppressed, as they are in most informal expositions, and the reader is asked to supply them wherever a sort of sorts is needed. The development is not mechanically checked, a fact that the essay repeats because it bears on how much of the apparatus may be trusted.

3.1 Terms and sorts

The terms are generated by the grammar

$$t \;::=\; x \;\mid\; \mathsf{Prop} \;\mid\; \mathsf{Type} \;\mid\; \Pi x{:}A.\, B \;\mid\; \lambda x{:}A.\, t \;\mid\; t\,u \;\mid\; c \;\mid\; \mathsf{match}\ t\ \mathsf{with}\ \overline{c\,\bar{x} \Rightarrow t_c},$$

where x ranges over variables, c over constructors of declared inductive types, and the sorts Prop and Type stand for the impredicative sort of propositions and the predicative hierarchy of types. A non-dependent product is written $A \to B$. A context Γ is a list of declarations $x{:}A$, and a closed term is a term typable in the empty context, which for the purposes of this essay is the context containing only the declared inductive types and their constructors. A judgment $\Gamma \vdash t : A$ is read as t has type A in Γ. The rules for variables, products, abstraction, application and conversion are the usual ones:

$$\frac{(x{:}A) \in \Gamma}{\Gamma \vdash x : A} \qquad \frac{\Gamma \vdash A : s_1 \quad \Gamma, x{:}A \vdash B : s_2}{\Gamma \vdash \Pi x{:}A.\, B : s_2} \ (s_2 = \mathsf{Prop})$$

$$\frac{\Gamma, x{:}A \vdash t : B \quad \Gamma \vdash \Pi x{:}A.\, B : s}{\Gamma \vdash \lambda x{:}A.\, t : \Pi x{:}A.\, B} \qquad \frac{\Gamma \vdash f : \Pi x{:}A.\, B \quad \Gamma \vdash a : A}{\Gamma \vdash f\, a : B[a/x]}$$

$$\frac{\Gamma \vdash t : A \quad A \equiv_{\beta\iota} A' \quad \Gamma \vdash A' : s}{\Gamma \vdash t : A'}$$

The first product rule is the rule of impredicativity: a product into Prop is again a proposition whatever the sort of its domain. Without it the calculus is a predicative type theory with a different expressive profile, and the proofs below that quantify over propositions (the definition of absurdity, in particular) would be unavailable. Conversion is $\beta\iota$-conversion, that is, β-reduction together with the reduction of a match applied to a constructor.

3.2 Metatheoretic facts that are cited and not proved

Three facts are used. The first is strong normalization: every well-typed term has no infinite reduction sequence, which Coquand and Huet established for the calculus of constructions and which extends to the inductive version [2, 9]. The second is that every well-typed term has a unique $\beta\iota$-normal form, by confluence. The third is the consequence of these two that the type $\bot := \Pi p{:}\mathsf{Prop}.\, p$ is uninhabited in the empty context, so that the theory is consistent relative to the metatheory in which the normalization argument is carried out. None of the three is proved in this essay. All three are standard, and the essay's use of them is limited to the arguments in Sections 7 and 14.

Lemma 3.1 (Shape of normal forms)**.** *Let Γ be a context and let t be a term in normal form with $\Gamma \vdash t : C$, where C is neither a product nor a sort. Then t is either a constructor applied to arguments or a* neutral *term, that is, a variable $x \in \Gamma$ applied to zero or more arguments.*

Proof. The cases are by the form of t. A normal term cannot be an abstraction, since an abstraction has a product type and conversion preserves the head form of a type up to $\beta\iota$-equality, so a type that is not a product is not convertible to one. It cannot be a sort or a product, since these have sorts as types. It cannot be a match on a constructor, since that would be a ι-redex. A match on a variable is permitted and is neutral in the extended sense that its head is a variable, and a match on a neutral term is likewise neutral. The remaining forms are a constructor application and an application $f\, a_1 \cdots a_k$, whose head f is not an abstraction, since otherwise the application would be a β-redex, and not a constructor with the full complement of arguments

unless the term is a constructor application. Hence the head is a variable or a constructor. □

The lemma is the whole of the proof-theoretic machinery that Section 7 needs. Its use there is to show that a certain proposition has no proof in a certain context, by arguing that any proof, being reducible to a normal form, would have to end in a variable that the context does not supply.

3.3 The object-level declarations

The objects that the essay is about are declared as parameters and inductive types. A parameter is a typed variable with no definition, and it is the formal counterpart of a concept the essay uses without defining it. The declarations are given in Listing 1 in the concrete syntax of a proof assistant. They are not intended to be loaded into one.

Listing 1: Object-level declarations (unchecked).

```
Parameter Work : Set.
Parameter Essay : Set.
Inductive Status : Set := Design | Pilot | Measured.
Inductive Relation : Set :=
  Transfer | Tension | Composition | Boundary | Triangulation.
Record Combination := { members : list Work; arity_ok : length members = 2 \/ length members
     = 3 }.
Record Brief := { comb : Combination; rel : Relation; roles : list Work }.
Parameter Develops : Brief -> Essay -> Prop.
Parameter Warranted : Essay -> Essay -> Prop.
Definition H (e1 e2 : Essay) : Prop := Warranted e1 e2.
```

Two features of the declarations matter. The predicate `Warranted` is a parameter. It is not defined from any other notion, and the only way to obtain a proof of it is to be given one. This is the formal expression of the informality of Definition 1.1. Second, the type `Develops` relates briefs to essays by a relation and not by a function, so that a brief may be developed into no essay, one essay or many. The distinction between a combination, a brief and an essay, which the first of the two essays insists on, is thereby carried by the types: a combination is a set of works, a brief is a combination with a relation and a role ordering, and an essay is whatever stands in the relation `Develops` to a brief.

Status of H. Unproved. The hypothesis is the proposition `Warranted e1 e2`, a parameter with no definition. Nothing in this section supplies an inhabitant.

4 A Process Algebra of Circles

The second formalism is a process algebra in the manner of SpherePOP, a calculus in which computation proceeds by popping nested circles. A circle holds an operator and a list of arguments, each of which is a value or another circle, and a pop replaces a circle whose arguments are all values by the result of applying its operator to them. The calculus is used here in a fragment adequate to the decisions at hand. It is not a faithful reproduction of any particular

implementation, and the lifecycle layer in Section 4.3 is a simplification of a fuller disposition calculus.

4.1 Circles and pops

Fix a set $\mathcal{V}$ of values, partitioned into sorts including booleans, naturals, essays and hashes, and a set $\mathcal{O}$ of operators, each with a signature $\mathrm{sig}(o) = (\sigma_1, \ldots, \sigma_n)$ and a partial result function $\delta_o : \mathcal{V}^n \rightharpoonup \mathcal{V}$.

Definition 4.1 (Circle)**.** A *circle* is either a value $v \in \mathcal{V}$ or an expression $\langle o\ c_1 \cdots c_n \rangle$ where $o \in \mathcal{O}$ and each c_i is a circle. The *size* $|c|$ of a circle is the number of non-value circles it contains, so $|v| = 0$ and $|\langle o\ c_1 \cdots c_n \rangle| = 1 + \sum_i |c_i|$.

An operator application is *admissible* if the argument values have the sorts the signature demands and δ_o is defined on them. The pop relation $c \rightsquigarrow c'$ is the least relation closed under two rules, in which v_i ranges over values:

$$\frac{\langle o\ v_1 \cdots v_n \rangle \text{ admissible}}{\langle o\ v_1 \cdots v_n \rangle \rightsquigarrow \delta_o(v_1, \ldots, v_n)}\ (\textsc{Pop}) \qquad \frac{c_i \rightsquigarrow c_i' \quad c_1, \ldots, c_{i-1} \text{ values}}{\langle o\ c_1 \cdots c_i \cdots c_n \rangle \rightsquigarrow \langle o\ c_1 \cdots c_i' \cdots c_n \rangle}\ (\textsc{Arg})$$

The second rule fixes the order: the leftmost argument that is not yet a value is the one that is reduced, so that a circle is popped from the inside out and from the left. A circle with no applicable rule is in *normal form*. A circle in normal form is either a value or stuck on an inadmissible application.

Lemma 4.2 (Determinism)**.** *If $c \rightsquigarrow c'$ and $c \rightsquigarrow c''$ then $c' = c''$.*

Proof. By induction on the derivation of the first step. If the first step is by Pop, then every argument of c is a value, so Arg does not apply, and the result is $\delta_o(v_1, \ldots, v_n)$ in both derivations since δ_o is a function. If the first step is by Arg at position i, then c_i is the leftmost non-value argument, so the second step is also by Arg at the same position, and the induction hypothesis applied to $c_i \rightsquigarrow c_i'$ and $c_i \rightsquigarrow c_i''$ gives $c_i' = c_i''$. □

Lemma 4.3 (Termination)**.** *Suppose every δ_o returns a value. Then every pop strictly decreases the size, and there is no infinite sequence $c_0 \rightsquigarrow c_1 \rightsquigarrow \cdots$.*

Proof. A step by Pop replaces a circle of size 1 (all arguments being values) by a value of size 0, so the size of the whole circle decreases by exactly one. A step by Arg replaces c_i by c_i' with $|c_i'| < |c_i|$ by the induction hypothesis, and the sizes of the other arguments are unchanged. Since the size is a natural number, no infinite sequence exists. □

Lemmas 4.2 and 4.3 together give that a circle has at most one normal form and reaches it in a number of pops bounded by its size. They also show why the calculus is a poor model of deliberation: a circle that always terminates cannot represent a decision that is deferred without limit. The lifecycle layer below is needed for that.

4.2 An example reduction

Let and be the ternary conjunction on booleans and let status, qualified and demo be unary operators on essays returning a boolean according to whether the essay carries a status line in its opening paragraph, whether its title qualifies the figure it states, and whether its claims about earlier code are accompanied by their demonstrations. Let e_1 and e_2 be the two essays and suppose, for the example, that e_1 has the first and third properties and lacks the second. Then

$$\begin{aligned}
&\langle \mathsf{and}\ \langle \mathsf{status}\ e_1\rangle\ \langle \mathsf{qualified}\ e_1\rangle\ \langle \mathsf{demo}\ e_1\rangle\rangle \\
&\quad \rightsquigarrow \langle \mathsf{and}\ \mathsf{tt}\ \langle \mathsf{qualified}\ e_1\rangle\ \langle \mathsf{demo}\ e_1\rangle\rangle \\
&\quad \rightsquigarrow \langle \mathsf{and}\ \mathsf{tt}\ \mathsf{ff}\ \langle \mathsf{demo}\ e_1\rangle\rangle \\
&\quad \rightsquigarrow \langle \mathsf{and}\ \mathsf{tt}\ \mathsf{ff}\ \mathsf{tt}\rangle \\
&\quad \rightsquigarrow \mathsf{ff}.
\end{aligned}$$

The circle has size four and reaches a value in four pops, as Lemma 4.3 requires. The value ff is the answer to a question about the essay's current text. It is not the answer to the question whether the essay is warranted, and the example is included in part to make the distinction visible: a circle can compute a conjunction of editorial conditions, and nothing in the calculus connects that conjunction to H. The connection is an axiom, introduced and labelled as such in Section 13.

4.3 The lifecycle of a commitment

Decisions are modelled as commitments that move through states. A commitment is a triple (c, s, h) of a circle c, a state s, and a history h that is a finite sequence of labels. The states are

$$\mathrm{Live} = \{\mathsf{open}, \mathsf{bound}, \mathsf{transformed}, \mathsf{verified}, \mathsf{vfailed}\}, \qquad \mathrm{Term} = \{\mathsf{collapsed}, \mathsf{refused}\},$$

and the labels are POP, BIND(p) for a principal p, TRANSFORM, VERIFY^{+}, VERIFY^{-}, COLLAPSE and REFUSE. A predicate auth(p) holds of the principals with authority to bind. The transition relation is generated by

$$\begin{aligned}
&(c, \mathsf{open}, h) \xrightarrow{\mathsf{BIND}(p)} (c, \mathsf{bound}, h \cdot \mathsf{BIND}(p)) && \text{if } \mathrm{auth}(p), \\
&(c, \mathsf{bound}, h) \xrightarrow{\mathsf{TRANSFORM}} (c', \mathsf{transformed}, h \cdot \mathsf{TRANSFORM}) \\
&(c, \mathsf{transformed}, h) \xrightarrow{\mathsf{VERIFY}^{+}} (c, \mathsf{verified}, h \cdot \mathsf{VERIFY}^{+}) \\
&(c, \mathsf{transformed}, h) \xrightarrow{\mathsf{VERIFY}^{-}} (c, \mathsf{vfailed}, h \cdot \mathsf{VERIFY}^{-}) \\
&(c, \mathsf{vfailed}, h) \xrightarrow{\mathsf{TRANSFORM}} (c', \mathsf{transformed}, h \cdot \mathsf{TRANSFORM}) \\
&(c, \mathsf{verified}, h) \xrightarrow{\mathsf{COLLAPSE}} (c, \mathsf{collapsed}, h \cdot \mathsf{COLLAPSE}) \\
&(c, s, h) \xrightarrow{\mathsf{REFUSE}} (c, \mathsf{refused}, h \cdot \mathsf{REFUSE}) && \text{if } s \in \mathrm{Live}.
\end{aligned}$$

A commitment starts as $(c, \mathsf{open}, [\mathsf{POP}])$. Two design decisions are visible in the rules. Refusal is available from every live state and not only from the first one, so that a decision to decline can be taken at any point of the process. A failed verification is a live state distinct from refusal, so that the verifier has no authority to terminate the commitment and can only report. A further transformation may follow a failed verification, which makes the pair transformed and vfailed a cycle.

Lemma 4.4 (Refusal is always available)**.** *For every commitment (c, s, h) with $s \in$* Live *there is a transition to* refused*.*

Proof. Immediate from the last rule. □

Lemma 4.5 (Terminal states are final)**.** *No transition has a source in* Term*.*

Proof. Every rule other than the last has a live state as its source, and the last rule requires $s \in$ Live. □

Theorem 4.6 (Publication safety)**.** *Let $(c, \mathsf{open}, [\mathsf{POP}]) \to^* (c', \mathsf{collapsed}, h)$. Then h contains a label $\mathsf{BIND}(p)$ with $\mathsf{auth}(p)$, a later label* TRANSFORM*, and a later label VERIFY^+ with no* TRANSFORM *between that VERIFY^+ and the final* COLLAPSE*, and h contains no* REFUSE*.*

Proof. Define an invariant $I(s, h)$ for each state as follows. For open, $h = [\mathsf{POP}]$. For bound, h begins with $\mathsf{POP} \cdot \mathsf{BIND}(p)$ for some p with $\mathsf{auth}(p)$. For transformed, h contains such a prefix and ends in TRANSFORM. For vfailed, h contains such a prefix and ends in VERIFY^- preceded by some TRANSFORM. For verified, h contains such a prefix and ends in $\mathsf{TRANSFORM} \cdot \mathsf{VERIFY}^+$. In every case h contains no REFUSE. The invariant holds initially and is preserved by each rule other than the last, by inspection of the rules; the last rule leads to refused, which is not collapsed, and by Lemma 4.5 cannot be left. Hence a sequence ending in collapsed uses only the first six rules, and its last transition is COLLAPSE from verified, at which point the invariant for verified yields the claim. □

Theorem 4.6 states what it takes, in the model, for a publication to occur: an authorized binding, a transformation, a successful verification of the transformed text, and no refusal. It does not state that the transformation was an improvement or that the verification was sound. The verification label records that a check was passed. What the check checks is outside the calculus.

4.4 The ledger and the timestamp

The argument for publishing now rests in part on the order of events in a public record. To state it, let a *ledger* be a finite sequence of entries, and let $L \sqsubseteq L'$ mean that L is a prefix of L'.

Definition 4.7 (Append-only ledger)**.** A ledger update is a step $L \mapsto L \cdot \epsilon$ for an entry ϵ. A history of updates is a sequence $L_0 \sqsubseteq L_1 \sqsubseteq \cdots$.

Lemma 4.8 (Monotone order)**.** *If an entry ϵ occurs at position i of L_j then it occurs at position i of L_k for every $k \geq j$. Consequently, if entry ϵ is at a position less than that of ϵ' in some L_j in which both occur, the same holds in every L_k, $k \geq j$.*

Proof. By induction on $k - j$. The base case is trivial, and the inductive step uses that $L_{k+1} = L_k \cdot \epsilon''$ has L_k as a prefix. □

Corollary 4.9 (A hash precedes later publication)**.** *If a tag entry recording a digest of the two essays is appended at a position before the position of a later* COLLAPSE *entry for their publication, then in every later ledger the tag precedes the publication.*

The corollary formalizes the claim, made on both sides of the earlier debate, that the order of events can be preserved by a tagged commit without wide publication. It does not formalize any claim about the value of the order, which is a matter for the author.

> STATUS OF H. Unproved. Theorem 4.6 shows what a publication would require in the calculus, and Corollary 4.9 shows that holding with a hash preserves order. Neither shows that a publication is warranted, and the calculus has no rule that would.

5 The Objects of the Two Essays

The first of the two essays rests on three claims that can be stated exactly: the count of combinations, the absence of any bound relating that count to a count of essays, and the scheme for addressing a combination by an integer. Each is a small theorem about finite sets, and each is proved here because the first essay was criticized, correctly, for stating the second of them in a form that a reader could take as a bound.

5.1 Counting combinations

Let W be a finite set of works with $|W| = n$, and write $\binom{n}{k}$ for the number of k-element subsets of W.

Lemma 5.1 (Increment of triples)**.** *For all $n \geq 0$, $\binom{n+1}{3} - \binom{n}{3} = \binom{n}{2}$.*

Proof. By Pascal's rule, $\binom{n+1}{3} = \binom{n}{3} + \binom{n}{2}$. Combinatorially, a 3-subset of $W \cup \{w\}$ either omits w, and is one of $\binom{n}{3}$ subsets of W, or contains w, and is determined by the 2-subset of W that accompanies it. □

Theorem 5.2 (Count of combinations)**.** *The number of unordered pairs and triples of W is $\binom{n}{2} + \binom{n}{3}$. This number is strictly increasing in n for $n \geq 2$, grows as the cube of n, and satisfies $\binom{n}{3} \geq \frac{1}{6}(n-2)^3$.*

Proof. The first claim is the definition of the binomial coefficient. Strict monotonicity follows from Lemma 5.1 and $\binom{n}{2} > 0$ for $n \geq 2$, together with $\binom{n+1}{2} > \binom{n}{2}$. For the lower bound, $\binom{n}{3} = n(n-1)(n-2)/6 \geq (n-2)^3/6$ because $n(n-1) \geq (n-2)^2$ for $n \geq 2$. The cubic upper bound $\binom{n}{3} \leq n^3/6$ is immediate. □

For the corpus snapshot examined in the first essay, grouped by a provisional rule into $n = 488$ works, evaluation gives $\binom{488}{2} = 118{,}828$ pairs and $\binom{488}{3} = 19{,}250{,}136$ triples, for $19{,}368{,}964$ combinations. In a proof assistant this evaluation is performed by computation on numerals, and its result is a closed term of an equality type. The numbers are exact, and they are exact *given* n. The sensitivity to n is the content of the next lemma.

Lemma 5.3 (Sensitivity to the grouping rule)**.** *Splitting one work into two raises the number of triples by* $\binom{n}{2}$*. Merging two works into one lowers it by* $\binom{n-1}{2}$*. For* $n = 488$ *the first change is* $118{,}828$ *and the second is* $118{,}341$*.*

Proof. The first claim is Lemma 5.1. For the second, $\binom{n}{3} - \binom{n-1}{3} = \binom{n-1}{2}$ is Lemma 5.1 at $n-1$. The values follow by evaluation. □

The lemma is the formal reason why the first essay describes the count as exact under a provisional grouping rule and declines to describe it as a measurement of the number of distinct works. A single misclassification of a draft as a separate work changes the count by about as many triples as the corpus contains pairs. The effect is a change of well under one percent, and the essay does not claim that it matters for the order of magnitude. The point is only that the exactness of the arithmetic is the exactness of a function, and a function is exact whatever its argument.

5.2 No bound on essays

The first essay was criticized for a sentence asserting that the number of essays obtainable from the corpus is bounded by the number of combinations. The sentence is false, and the following theorem says in what way.

Let a *development structure* be a tuple $S = (W, B, E, D)$ in which W is a finite set of works, B is the finite set of briefs over W (a brief being a combination together with a relation and an ordering of its members), E is a set of essays, and $D \subseteq B \times E$ is the relation of development. Let $\mathrm{ess}(S) = \{e \in E \mid \exists b.\, (b, e) \in D\}$ be the set of developed essays.

Theorem 5.4 (No bound)**.** *For every* $n \geq 3$ *and every* $m \in \mathbb{N}$ *there is a development structure* S *with* $|W| = n$ *and* $|\mathrm{ess}(S)| = m$*. In particular* $|\mathrm{ess}(S)|$ *may be zero, may be less than the number of combinations, and may exceed the number of briefs.*

Proof. Fix W with $|W| = n$ and let B be the set of briefs over W, which is nonempty since $n \geq 3$. Choose one brief $b_0 \in B$. Let $E = \{e_1, \ldots, e_m\}$ be any set of m distinct essays and let $D = \{(b_0, e_i) \mid 1 \leq i \leq m\}$. Then $\mathrm{ess}(S) = E$ and $|\mathrm{ess}(S)| = m$. The case $m = 0$ gives $D = \emptyset$. Since m is arbitrary, it may be chosen below $\binom{n}{3}$ or above $|B|$. □

Remark 5.5. The theorem is trivial, and its triviality is the point. A proof that the count of combinations constrains the number of essays would need a hypothesis tying essays to briefs, such as that every essay is developed from exactly one brief and every brief yields at most one essay, and the first essay does not assume that. A reader who adds the hypothesis obtains the bound $|\mathrm{ess}(S)| \leq |B|$ and so a bound by the number of briefs, which exceeds the number

of combinations by a small constant factor. The hypothesis is not supported by anything in the first essay, and the second essay's staged pipeline in fact permits a brief to be revised into several candidate developments.

The number of briefs over W does have an upper bound under the generator's conventions. There are four relations available to a pair and five to a triple (the fifth being triangulation), and the roles may be ordered in 2! or 3! ways.

Lemma 5.6 (Briefs per combination)**.** *Under the generator's conventions the number of briefs satisfies* $|B| \le 8\binom{n}{2} + 30\binom{n}{3}$.

Proof. A pair supports at most $4 \cdot 2! = 8$ briefs and a triple at most $5 \cdot 3! = 30$. Summing over the combinations gives the bound. □

The bound is generous, because many role orderings are equivalent under some relations. It shows that the multiplication of briefs over combinations is by a constant and not by a function of n, and the first essay's remark that relations and orderings multiply briefs severalfold without adding a combination is therefore correct.

Non-Result 5.7 (The yield). This essay does not prove that the generator yields developed essays at any positive rate, that is, that there exists $\rho > 0$ such that a randomly drawn brief is developed into an essay with probability at least ρ. The statement is neither assumed nor refuted. Theorem 5.4 shows that the structure of the objects permits every rate, including zero, and the rate is the quantity that the study proposed in the first essay would measure.

5.3 Addressing combinations by integer

The first essay proposes naming each triple by an integer through the combinatorial number system. The scheme is classical [5]. It is proved here in a form that makes the append-only requirement a theorem and not a recommendation.

Write a 3-subset of $\mathbb{N}$ as $c_1 < c_2 < c_3$ and define its rank

$$\rho_3(c_1, c_2, c_3) = \binom{c_1}{1} + \binom{c_2}{2} + \binom{c_3}{3}.$$

More generally, for $k \ge 1$ and $c_1 < \cdots < c_k$ define $\rho_k = \sum_{i=1}^{k} \binom{c_i}{i}$.

Theorem 5.8 (Bijection)**.** *For every $k \ge 1$ and $n \ge k$ the map ρ_k is a bijection from the k-subsets of $\{0, \dots, n-1\}$ onto $\{0, 1, \dots, \binom{n}{k} - 1\}$.*

Proof. By induction on k. For $k = 1$, $\rho_1(c) = c$ and the claim is immediate. Suppose it holds for $k-1$. Partition the k-subsets of $\{0, \dots, n-1\}$ by their largest element $c_k = c$, for $c = k-1, \dots, n-1$. The subsets with largest element c correspond to the $(k-1)$-subsets of $\{0, \dots, c-1\}$, of which there are $\binom{c}{k-1}$, and their ranks are $\binom{c}{k} + \rho_{k-1}(c_1, \dots, c_{k-1})$. By the induction hypothesis the values ρ_{k-1} range bijectively over $\{0, \dots, \binom{c}{k-1} - 1\}$, so the ranks of this block fill the interval $[\binom{c}{k}, \binom{c}{k} + \binom{c}{k-1}) = [\binom{c}{k}, \binom{c+1}{k})$ by Pascal's rule. These intervals are pairwise

disjoint for distinct c and their union over $c = k-1, \dots, n-1$ is $[0, \binom{n}{k})$, since $\binom{k-1}{k} = 0$ and the right endpoint of the last interval is $\binom{n}{k}$. Hence ρ_k is a bijection onto $\{0, \dots, \binom{n}{k} - 1\}$. □

Corollary 5.9 (Recovery)**.** *The subset of rank r is recovered greedily: c_k is the largest c with $\binom{c}{k} \le r$, and the remaining elements are the subset of rank $r - \binom{c_k}{k}$ with arity $k-1$.*

The ranking is invertible in time linear in k times the search for each c_i. A combination can therefore be named, stored as a number, and regenerated without being written down. This is the sense in which any of the nineteen million triples can be addressed, and it is a statement about addresses and not about essays.

Lemma 5.10 (Append-only stability)**.** *Let $W_n = \{0, \dots, n-1\}$ and $W_{n+1} = W_n \cup \{n\}$. For every k-subset s of W_n, the rank $\rho_k(s)$ is the same whether s is regarded as a subset of W_n or of W_{n+1}. The ranks of the new subsets, those containing n, are exactly the integers in $[\binom{n}{k}, \binom{n+1}{k})$.*

Proof. The formula for ρ_k does not mention n, so the first claim is immediate. The new subsets have largest element n, and the second claim is the block computation in the proof of Theorem 5.8. □

Proposition 5.11 (Insertion breaks addresses)**.** *Suppose the works are indexed by a fixed total order and a new work is inserted at position p, shifting the indices of all works at positions $\ge p$ up by one. Then the rank of every triple containing a work at a position $\ge p$ strictly increases.*

Proof. On the domain $c_1 < c_2 < c_3$ of nonnegative integers one has $c_i \ge i - 1$, and the function $c \mapsto \binom{c}{i}$ is strictly increasing on $c \ge i - 1$. A shift increases each affected coordinate by one and leaves the others unchanged, so each affected term strictly increases and the others are unchanged. The shift preserves the strict order of the coordinates, so the shifted triple remains in the domain. Summing, the rank strictly increases when at least one coordinate is affected. □

The proposition is the formal content of the first essay's observation that sorting the works by title makes the address of a triple depend on the works that happen to sort before it. Under an append-only ordering the addresses are stable by Lemma 5.10, and the ranks of the new triples occupy a fresh interval. The lemma is silent on a further problem that the first essay raises, which is that grouping is heuristic.

Lemma 5.12 (Retirement leaves gaps)**.** *Suppose a group with index $j < n$ is retired after a merge, and the remaining groups keep their indices. Then the set of ranks of triples of live groups is a proper subset of $\{0, \dots, \binom{n}{3} - 1\}$ whenever $n \ge 4$.*

Proof. There is at least one triple of W_n containing j, and its rank is not the rank of any triple of live groups since ρ_3 is injective by Theorem 5.8. That rank is in the interval, so the interval is not covered. The assumption $n \ge 4$ ensures that at least one live triple exists. □

Retirement without renumbering therefore costs the property that addresses form an initial segment. The cost is paid in bookkeeping: an enumeration by rank must skip retired values,

and the registry must record which indices are retired and what replaced them. The alternative of renumbering restores the segment and breaks the stability of Lemma 5.10. No scheme has both properties once groups merge, which is a consequence of the heuristic nature of the grouping and not of the numbering.

Status of H. Unproved. Theorems 5.2 through 5.8 are proved, and none bears on H. Theorem 5.4 shows that the combination count permits every number of essays. Non-Result 5.7 records that the rate at which briefs become essays is not established.

6 The Rebound Effect

The generator described in the first essay makes a brief cheap to produce. The statement that a brief is cheap is a statement about one component of its cost. The other component, the cost of reading the brief with enough care to decide whether it supports an argument, is borne by the author and is not reduced by the generator at all. The theorems of this section make this distinction exact, and they show that the effect that Jevons described for coal [4] appears in two forms, one of which depends on a parameter this essay does not know and the other of which does not.

6.1 The model

Let $p > 0$ be the *effective price* of a brief, measured in units of the author's time, and let $q(p)$ be the number of briefs demanded per period at that price. The demand is a decreasing function of price, which the essay takes as an assumption and not as an observation. The price is the sum of two parts, $p = g + v$, where $g > 0$ is the cost of generating the brief and $v > 0$ is the cost of reviewing it. The total effort expended per period is $E(p) = p\,q(p)$, the review burden is $V(p) = v\,q(p)$, and the generation burden is $G(p) = g\,q(p)$, so that $E = G + V$.

Assumption 6.1 (Constant elasticity)**.** For $p > 0$ the demand is $q(p) = A\,p^{-\varepsilon}$ with $A > 0$ and $\varepsilon > 0$.

The assumption is made because it permits closed forms. It is not defended. Real demand for briefs is presumably not of this form over any large range of prices, and the essay returns to this in Section 6.5.

For a general differentiable decreasing demand the point elasticity is $\eta(p) = -p\,q'(p)/q(p)$, which is positive, and Assumption 6.1 is the case $\eta \equiv \varepsilon$.

6.2 The threshold

Proposition 6.2 (Local rebound)**.** *For differentiable q, $E'(p) = q(p)\,(1 - \eta(p))$. Hence, at a price where $\eta(p) > 1$, a small reduction in price strictly increases total effort, and where $\eta(p) < 1$ it strictly decreases total effort.*

Proof. By the product rule, $E'(p) = q(p) + p\,q'(p) = q(p)\,(1 + p\,q'(p)/q(p)) = q(p)(1 - \eta(p))$. Since $q(p) > 0$, the sign of E' is the sign of $1 - \eta$. A small reduction in p changes E by approximately $-E'(p)\,\Delta p$, which is positive exactly when $E'(p) < 0$, that is, when $\eta(p) > 1$. □

Theorem 6.3 (Jevons threshold)**.** *Under Assumption 6.1, for any $0 < p' < p$ one has $E(p') > E(p)$ if $\varepsilon > 1$, $E(p') = E(p)$ if $\varepsilon = 1$, and $E(p') < E(p)$ if $\varepsilon < 1$.*

Proof. Here $E(p) = A\,p^{1-\varepsilon}$. If $\varepsilon > 1$ the exponent $1 - \varepsilon$ is negative, the map $x \mapsto x^{1-\varepsilon}$ is strictly decreasing on $(0, \infty)$, and $p' < p$ gives $E(p') > E(p)$. If $\varepsilon = 1$, $E \equiv A$. If $\varepsilon < 1$ the exponent is positive, the map is strictly increasing, and $p' < p$ gives $E(p') < E(p)$. □

Theorem 6.3 is the Jevons paradox in its textbook form: when the demand is elastic, a fall in price raises total expenditure of the resource, here the author's effort. The essay does not know whether the demand for briefs is elastic. The question is empirical and the corpus does not answer it.

Non-Result 6.4 (Elasticity). This essay does not prove that $\varepsilon > 1$, nor that $\varepsilon \le 1$, nor that the demand is of the form in Assumption 6.1.

6.3 The decomposition

The more informative result splits the price. A generator reduces g and leaves v alone. Let $\lambda \in (0, 1)$ be the factor by which the generator's improvement reduces the generation cost, so that the price falls from $g + v$ to $\lambda g + v$.

Theorem 6.5 (Review burden always rises)**.** *Under Assumption 6.1, and for every $\lambda \in (0, 1)$ and every $\varepsilon > 0$:*

(*i*) *the demand rises by the factor $\kappa = \left(\dfrac{g + v}{\lambda g + v}\right)^{\varepsilon}$, and $1 < \kappa < \lambda^{-\varepsilon}$;*

(*ii*) *the review burden V rises by the same factor κ;*

(*iii*) *the total effort E changes by the factor $\kappa^{(\varepsilon-1)/\varepsilon}$, which exceeds one if and only if $\varepsilon > 1$;*

(*iv*) *as $v/g \to \infty$ the factor $\kappa \to 1$, and as $v/g \to 0$ the factor $\kappa \to \lambda^{-\varepsilon}$.*

Proof. (i) Let $r = (g + v)/(\lambda g + v)$. Since $\lambda g + v < g + v$ one has $r > 1$, and $\kappa = r^{\varepsilon} > 1$. Moreover $\lambda(g + v) = \lambda g + \lambda v < \lambda g + v$, so $r < 1/\lambda$ and $\kappa < \lambda^{-\varepsilon}$. The demand at the lower price divided by the demand at the higher price is $(\lambda g + v)^{-\varepsilon}/(g + v)^{-\varepsilon} = r^{\varepsilon} = \kappa$.

(ii) The review cost per brief v is unchanged, so $V = vq$ changes by the factor by which q changes.

(iii) By Theorem 6.3's computation, $E = A\,p^{1-\varepsilon}$, so E changes by the factor $(\lambda g + v)^{1-\varepsilon}/(g + v)^{1-\varepsilon} = r^{\varepsilon-1} = \kappa^{(\varepsilon-1)/\varepsilon}$. Since $r > 1$, this exceeds one if and only if $\varepsilon - 1 > 0$.

(iv) Write $r = (1 + x)/(\lambda + x)$ with $x = v/g$. As $x \to \infty$, $r \to 1$. As $x \to 0$, $r \to 1/\lambda$. Taking the ε-th power gives the claim. □

Part (ii) is the result that matters for the question of publication. It holds without any condition on ε beyond positivity. Whatever the elasticity, a generator that cheapens the production of briefs and does nothing for their review raises the burden on the reviewer. The degree of the rise depends on how much of the price was generation, which is part (iv): if review dominates the price, the generator changes little, and if generation dominated, the generator changes a great deal. For the generator described in the first essay, which sends a prompt to a local model and receives an outline, the second case is the more plausible one, though the essay makes no estimate of the ratio v/g and states none.

Corollary 6.6 (Review is the scarce resource)**.** *If the author's review capacity per period is a fixed quantity s in units of briefs, then any reduction of g that raises the demand above s produces briefs that cannot be reviewed within the period.*

6.4 Backlog

The corollary is made exact by a recurrence. Let Q_t be the number of unreviewed briefs at the start of period t, let a be the arrival rate per period and s the review capacity per period.

Lemma 6.7 (Backlog growth)**.** *Suppose $Q_{t+1} = \max(0, Q_t + a - s)$ with $a, s \geq 0$ and $Q_0 \geq 0$. If $a > s$ then $Q_t = Q_0 + t(a - s)$ for all t, so the backlog is unbounded. If $a \leq s$ and $Q_0 = 0$ then $Q_t = 0$ for all t.*

Proof. If $a > s$, then $Q_t + a - s > 0$ for every $Q_t \geq 0$, so the maximum is the second argument, and induction on t gives the formula. If $a \leq s$ and $Q_t = 0$, then $Q_t + a - s \leq 0$ and $Q_{t+1} = 0$; induction gives the second claim. □

Theorem 6.8 (An author at capacity is pushed over by any improvement)**.** *Suppose that at the baseline price $p = g + v$ the arrival rate equals the review capacity, $q(g + v) = s$. Then for every $\lambda \in (0, 1)$ the arrival rate at the price $\lambda g + v$ is $\kappa s > s$, and the backlog grows by exactly $(\kappa - 1)s$ briefs per period, without bound.*

Proof. By Theorem 6.5(i), $a = \kappa s$ with $\kappa > 1$. Lemma 6.7 gives $Q_t = Q_0 + t(\kappa s - s)$. □

The theorem does not assert that the author is at capacity. It asserts that an author who is at capacity and who adopts a generator that reduces g will accumulate unreviewed briefs at a linear rate, whatever the size of the reduction, as long as there is one. The qualification matters, and it reappears in the next lemma, which concerns the pool of sources.

Lemma 6.9 (Provenance fraction)**.** *Suppose every generated brief is added to the pool from which sources are drawn, and that at most s briefs are reviewed per period. After t periods, with $a > s$ briefs generated per period, the fraction of the generated briefs that have been reviewed is at most s/a.*

Proof. At most st briefs have been reviewed and exactly at have been generated. The ratio is at most $st/(at) = s/a$. □

The lemma is the formal justification for the rule, stated in both essays, that generated briefs remain outside the source pool. If they were admitted, the fraction of the pool whose status had been established by a reader would be bounded above by s/a, which is strictly less than one and which tends to a smaller value as the generator improves. The rule is not a courtesy to the reader. It is what keeps the status of each source in the pool a fact about the source and not a fact about the throughput of the author.

6.5 Saturation and the limit of zero price

Assumption 6.1 has an implausible consequence: as $p \to 0$ the demand grows without bound. Real demand for briefs is bounded by what an author could ever want to read. Replace the demand by a saturating one, $q(p) = \min(A\,p^{-\varepsilon}, Q_{\max})$ for a finite $Q_{\max} > 0$.

Theorem 6.10 (Rebound vanishes at zero price)**.** *Let $p^* = (A/Q_{\max})^{1/\varepsilon}$. For $p \geq p^*$ the demand is $A\,p^{-\varepsilon}$, and for $0 < p \leq p^*$ it is $Q_{\max}$. Consequently $E(p) = p\,Q_{\max}$ for $p \leq p^*$, so E is strictly increasing in p on $(0, p^*]$ and $E(p) \to 0$ as $p \to 0^+$. For $p \geq p^*$ the conclusions of Theorem 6.3 hold.*

Proof. The expression $A\,p^{-\varepsilon}$ is decreasing in p and equals $Q_{\max}$ exactly at p^*, so it is at most $Q_{\max}$ for $p \geq p^*$ and at least $Q_{\max}$ for $p \leq p^*$. The minimum is therefore as stated. On $(0, p^*]$, $E(p) = pQ_{\max}$, which is linear and strictly increasing with limit zero. On $[p^*, \infty)$ the demand coincides with that of Assumption 6.1. □

The theorem says that a rebound can exist at moderate prices and vanish at low ones: past a point, a further fall in price reduces the total expenditure, because the volume has stopped growing. The result is of independent interest for the thesis that storage and compute should be free, since it identifies the mechanism by which a free resource can fail to be consumed without limit, namely that the demand for it saturates. It is important to say what it does not say. The volume of briefs stays at $Q_{\max}$ and the review burden at $vQ_{\max}$, which is not small if $Q_{\max}$ is large. A reduction in total expenditure in the sense of E is a reduction in generation expenditure, while the review burden persists.

Non-Result 6.11 (Saturation). This essay does not prove that the demand for briefs saturates, nor that $Q_{\max} \leq s/v$, which is the condition for the review burden at saturation to be within the author's capacity. Without the second condition Theorem 6.8 applies at saturation as well, and the backlog grows.

6.6 Reading the section

The section proves three things and leaves three open. It proves that a generator which lowers only the generation cost raises the review burden for every positive elasticity (Theorem 6.5), that an author at capacity accumulates a linearly growing backlog (Theorem 6.8), and that the rebound in total effort disappears in the limit of zero price if the demand saturates (Theorem 6.10). It leaves open the elasticity, the saturation level, and the position of the author relative to capacity. Since the three open quantities decide whether the rebound argues for or against publication, the section decides nothing about publication.

Publishing the essays tends to raise the demand for the procedure they describe, and this is relevant. If a reader other than the author runs the generator, the briefs it produces are reviewed by that reader and do not enter the author's backlog; the rebound then bears on a different reviewer. Whether that is a benefit or a hazard depends on whether the readers are the persons whom the author wishes to burden. The essay does not take a position.

Status of H. Unproved. Theorem 6.5 holds for every positive elasticity and every efficiency gain, and it shows that review burden rises. The same theorem holds whether H is true or false. Non-Results 6.4 and 6.11 record the parameters that would be needed to say more.

7 Proxies and Their Gaming

The two essays are unusual among descriptions of writing systems in that they refuse to accept a numerical score as evidence that a section is good. The refusal has a formal content, and this section supplies it. The common form of the difficulty was stated by Goodhart in 1975 as a remark about monetary aggregates [3], by Campbell in 1979 about social indicators [1], and by Strathern in 1997 in the form that a measure which becomes a target ceases to be a good measure [10]. Manheim and Garrabrant distinguish several mechanisms by which the failure occurs [7], and the three that matter here are the extremal, the regressional and the adversarial. Each has a theorem below.

7.1 Extremal gaming

Let an agent allocate a budget $B > 0$ between two activities. The amount $a \geq 0$ is spent on genuinely improving the object, and the amount $b \geq 0$ is spent on improving its appearance under the metric. The *target* is $t(a, b) = a$ and the *proxy* is $m(a, b) = a + \gamma b$, where $\gamma > 0$ is the rate at which expenditure on gaming raises the proxy. The feasible set is $F = \{(a, b) \in \mathbb{R}^2_{\geq 0} \mid a + b \leq B\}$.

Theorem 7.1 (Extremal Goodhart)**.** *If $\gamma > 1$, the proxy m has a unique maximizer on F, namely $(0, B)$, at which $t = 0$, while the target t has the unique maximizer $(B, 0)$, at which $m = B < \gamma B$. If $\gamma < 1$, both are uniquely maximized at $(B, 0)$. If $\gamma = 1$, every point of the segment $a + b = B$ maximizes m, and the target takes every value in $[0, B]$ on it.*

Proof. For $\gamma > 1$ and $(a, b) \in F$, $m = \gamma(a+b) - (\gamma-1)a \leq \gamma B - (\gamma-1)a \leq \gamma B$. Equality holds throughout if and only if $a+b = B$ and $a = 0$, that is, at $(0, B)$, so the maximizer is unique and $t(0, B) = 0$. The target a is at most B on F with equality only at $(B, 0)$, where $m = B < \gamma B$. For $\gamma < 1$ one has $m = a+\gamma b \leq a+\gamma(B-a) = \gamma B+(1-\gamma)a \leq B$, with equality if and only if $a = B$ (and $b = 0$). For $\gamma = 1$, $m = a + b \leq B$ with equality on the segment, and $t = a$ ranges over $[0, B]$ there. □

The theorem is elementary. Its force is in how little it assumes: all that is needed is that the gaming channel purchases proxy at a better rate than the genuine channel does, and then the

agent that maximizes the proxy spends nothing on the genuine channel. It makes no assumption that the agent is adversarial, dishonest or even aware of the distinction. A procedure that selects whatever scores best on the proxy will find the gaming channel by the same mechanism by which it finds anything else.

7.2 The entropy score

The earlier record of the pipeline's design proposed to accept a section when a blended score exceeded a threshold. The blend combined a model's numerical evaluation with a transformation of the unigram entropy of the text. The theorem of this subsection shows exactly which texts maximize the transformed entropy.

Let a text be a finite sequence $w = (w_1, \dots, w_n)$ of tokens with $n \geq 1$. Let c_x be the number of positions at which the token x occurs, let $p_x = c_x / n$, and let

$$H(w) = -\sum_{x} p_x \ln p_x,$$

where the sum is over the tokens that occur. This is the unigram entropy in nats. The score derived from it is $\sigma(w) = \min\bigl(100, \max(0,\ 100 - 10\,H(w))\bigr)$, and the blend with a model score $\mu \in [0, 100]$ is $\beta(w) = \alpha\mu + (1-\alpha)\sigma(w)$ for a weight $\alpha \in [0, 1)$.

Theorem 7.2 (Properties of the entropy score)**.** (*i*) *H is unchanged by every permutation of the positions of w.*

(*ii*) $0 \leq H(w) \leq \ln n$*, with $H(w) = 0$ if and only if all tokens of w are equal.*

(*iii*) $\sigma(w) \leq 100$*, with equality if and only if all tokens of w are equal.*

(*iv*) *For fixed μ and fixed n, the blend β is maximized exactly by the constant texts, and it is strictly smaller on every text with at least two distinct tokens.*

Proof. (i) The counts c_x depend only on the multiset of tokens, which a permutation of positions does not change, so H is unchanged.

(ii) Each term $-p_x \ln p_x$ is nonnegative since $0 < p_x \leq 1$, and it vanishes if and only if $p_x = 1$. Hence $H = 0$ if and only if a single token has probability one. For the upper bound, by Jensen's inequality applied to the concave function ln, $H = \sum p_x \ln(1/p_x) \leq \ln \sum p_x (1/p_x) = \ln d$, where d is the number of distinct tokens, and $d \leq n$.

(iii) Since $H \geq 0$, $100 - 10H \leq 100$ with equality if and only if $H = 0$, and the clamps preserve the claim. By (ii), $H = 0$ if and only if all tokens are equal.

(iv) Since $1 - \alpha > 0$, β is a strictly increasing function of σ for fixed μ. By (iii), σ attains its maximum 100 exactly on constant texts and is strictly less on all others. □

Part (i) says that the score cannot distinguish a text from any rearrangement of it, so a coherent paragraph and a word salad of the same words receive the same entropy and the same derived score. A sentence of thirteen tokens and a shuffled copy both have entropy of

about 1.925 nats, a fact that can be confirmed by direct computation and that is a special case of the theorem. Part (iv) says more: the transformation was designed, according to the text that accompanied it, to penalize repetition, and it does the opposite. The text that maximizes the blend for any model score is a single token repeated n times. An agent optimizing the blend by the extremal mechanism of Theorem 7.1 would be driven toward that text, to the extent that the model score did not resist.

Remark 7.3. It may be objected that no realistic agent would produce a constant text, since the model score would collapse. The objection is correct and is an argument for the model score and not for the entropy term: the entropy term contributes nothing to the discrimination between good and bad texts and contributes a bias toward repetition. A weighted sum of a useful term and a harmful one is not improved by the second, and the essays' decision to demote entropy to a diagnostic is the correct response.

7.3 Regressional gaming: selection overstates

The second mechanism does not require an optimizer with a strategy. It arises whenever items are selected because they score well on a proxy that is noisy. Suppose the proxy for a brief is $m = t + g$, where t is its true quality and g is a component that is independent of t and that has mean zero. The component g may be the gaming component, the idiosyncrasy of a model's self-assessment, or ordinary measurement noise; the argument does not distinguish them.

Lemma 7.4 (Association)**.** *Let g be an integrable real random variable and $\varphi : \mathbb{R} \to [0, 1]$ a nondecreasing function. Then $\mathbb{E}[g\,\varphi(g)] \geq \mathbb{E}[g]\,\mathbb{E}[\varphi(g)]$, with strict inequality if $\varphi(g)$ is not almost surely constant and g is not almost surely constant.*

Proof. Let g' be an independent copy of g. Because φ is nondecreasing, $(g - g')(\varphi(g) - \varphi(g')) \geq 0$ pointwise: if $g > g'$ then $\varphi(g) \geq \varphi(g')$, if $g < g'$ then $\varphi(g) \leq \varphi(g')$, and if $g = g'$ the product is zero. Taking expectations, which are finite since $|g\varphi(g)| \leq |g|$, gives $2\mathbb{E}[g\varphi(g)] - 2\mathbb{E}[g]\mathbb{E}[\varphi(g)] \geq 0$. The product is strictly positive on the event $g \neq g',\ \varphi(g) \neq \varphi(g')$, which has positive probability when $\varphi(g)$ is not almost surely constant. □

Theorem 7.5 (Selection overstates the target)**.** *Let t and g be independent real random variables with g integrable and $\mathbb{E}[g] = 0$, let $m = t + g$, and let c be a threshold with $\mathbb{P}(m \geq c) > 0$. Then*

$$\mathbb{E}[m \mid m \geq c] - \mathbb{E}[t \mid m \geq c] = \mathbb{E}[g \mid m \geq c] \geq 0.$$

The inequality is strict if g is not almost surely constant and the function $\varphi(x) = \mathbb{P}(t \geq c - x)$ is not almost surely constant under the law of g.

Proof. The equality is the linearity of conditional expectation. For the inequality, by independence and Fubini,

$$\mathbb{E}[g\,\mathbf{1}\{m \geq c\}] = \mathbb{E}\big[g\,\mathbb{P}(t \geq c - g \mid g)\big] = \mathbb{E}[g\,\varphi(g)],$$

where $\varphi(x) = \mathbb{P}(t \geq c - x)$ is nondecreasing in x and takes values in $[0, 1]$. Lemma 7.4 and $\mathbb{E}[g] = 0$ give $\mathbb{E}[g\varphi(g)] \geq 0$, and strictly so under the stated condition. Dividing by $\mathbb{P}(m \geq c) > 0$ gives the claim. □

Theorem 7.5 says that the briefs which pass a proxy filter have, on average, a proxy value higher than their true value, by an amount that is positive whenever the proxy has independent noise at all. No assumption is made about the distribution of the noise beyond independence and zero mean. The implication for the study proposed in the first essay is direct: if briefs are selected for examination because a model or a heuristic scored them highly, then the examination will find the selected briefs less good than the scores indicated, on average, and the shortfall is not evidence of malfunction. A pre-specified sample drawn without regard to the score avoids the effect, which is one reason the first essay insists that the draws be fixed before the judgments are made.

7.4 The best of n

A third way in which more generation can look like more quality is by selection among many.

Theorem 7.6 (Expected maximum)**.** *Let $X_1, X_2, \ldots$ be independent identically distributed integrable random variables and $M_n = \max_{i \leq n} X_i$. Then $\mathbb{E}[M_{n+1}] \geq \mathbb{E}[M_n]$, and the inequality is strict if X_1 is not almost surely constant.*

Proof. Since $M_{n+1} = \max(M_n, X_{n+1}) \geq M_n$ pointwise, $\mathbb{E}[M_{n+1}] \geq \mathbb{E}[M_n]$. If X_1 is not almost surely constant there is u with $\mathbb{P}(X_1 \leq u) > 0$ and $\mathbb{P}(X_1 > u) > 0$. Then $\mathbb{P}(M_n \leq u,\ X_{n+1} > u) = \mathbb{P}(X_1 \leq u)^n \mathbb{P}(X_1 > u) > 0$, and on this event $M_{n+1} > M_n$, so the inequality is strict. □

Proposition 7.7 (Apparent quality without real quality)**.** *There is a distribution of briefs for which the true quality t of every brief equals a constant τ and the proxy $m = \tau + g$ has a nonconstant noise component g, such that the expected proxy value of the best-scoring of n drawn briefs is strictly increasing in n, while the true quality of the selected brief equals τ for every n.*

Proof. Take $t \equiv \tau$ and g any nonconstant integrable variable. Then $\mathbb{E}[\max_{i \leq n} m_i] = \tau + \mathbb{E}[\max_{i \leq n} g_i]$, strictly increasing by Theorem 7.6, while $t = \tau$ for every brief, selected or not. □

The proposition is an existence statement, and its use is limited to showing that the improvement of the best-looking brief with the number of draws is compatible with no improvement at all in the briefs. It does not say that quality does not improve with more draws in the actual procedure.

7.5 Adversarial gaming: the ledger

The third mechanism is the one that the essays' own replacement for the numerical score invites. The ledger of obligations records, for each section, whether each of four obligations is discharged. A model may propose that an obligation is discharged, and a reader confirms or rejects the proposal. The design intends the confirmation to be the only route to a discharged

obligation. The question is whether the formal system can state this intention in a form that excludes the model's proposals from counting as confirmations, and the answer is that it can, using the normal-form argument of Lemma 3.1.

Let Γ_V be the context

$$\begin{aligned}&\mathsf{Obl} : \mathsf{Type},\\ &\mathsf{Asserted},\ \mathsf{ReaderOK},\ \mathsf{Confirmed} : \mathsf{Obl} \to \mathsf{Prop},\\ &\mathsf{model} : \Pi o{:}\mathsf{Obl}.\ \mathsf{Asserted}\ o,\\ &\mathsf{conf} : \Pi o{:}\mathsf{Obl}.\ \mathsf{Asserted}\ o \to \mathsf{ReaderOK}\ o \to \mathsf{Confirmed}\ o,\end{aligned}$$

in which the model is granted the strongest possible power: it can assert every obligation. The variable conf is the sole means of producing a confirmation, and it requires both the model's assertion and a reader's acceptance.

Theorem 7.8 (Self-confirmation is impossible)**.** *Let o be any term with $\Gamma_V \vdash o : \mathsf{Obl}$. There is no term t with $\Gamma_V \vdash t : \mathsf{Confirmed}\ o$. On the other hand, in the extended context $\Gamma_V,\ \mathsf{rok} : \Pi o{:}\mathsf{Obl}.\ \mathsf{ReaderOK}\ o$, the term $\lambda o.\ \mathsf{conf}\ o\ (\mathsf{model}\ o)\ (\mathsf{rok}\ o)$ has type $\Pi o{:}\mathsf{Obl}.\ \mathsf{Confirmed}\ o$.*

Proof. Suppose $\Gamma_V \vdash t : \mathsf{Confirmed}\ o$. By strong normalization we may take t in normal form. The type $\mathsf{Confirmed}\ o$ is neither a product nor a sort, and Confirmed is a variable and not an inductive type, so there is no constructor of that type. By Lemma 3.1, t is neutral: a variable of Γ_V applied to arguments. The only variable whose type ends in Confirmed is conf, since model ends in Asserted and the remaining variables are not functions into Confirmed. Hence $t = \mathsf{conf}\ o'\ a\ r$ with r of type $\mathsf{ReaderOK}\ o'$ in Γ_V. Take r in normal form. It is neutral by the same argument, headed by a variable whose type ends in ReaderOK. There is none: ReaderOK occurs in Γ_V only as the type of an argument of conf. This is a contradiction. The second claim is checked by the application rule: $\mathsf{model}\ o : \mathsf{Asserted}\ o$, $\mathsf{rok}\ o : \mathsf{ReaderOK}\ o$, and so $\mathsf{conf}\ o\ (\mathsf{model}\ o)\ (\mathsf{rok}\ o) : \mathsf{Confirmed}\ o$. □

The theorem formalizes the distinction on which the ledger rests. The model's assertion is derivable for every obligation and costs nothing, and the confirmation is derivable only when a reader supplies it. The extremal theorem applies with the gaming rate γ unbounded: a metric that counted asserted obligations would be maximized by asserting all of them and would convey no information. A metric that counts confirmed obligations is not gameable by the model, but it is gameable by the reader, and the theorem does not and cannot say anything about the diligence with which the term rok is produced.

Remark 7.9 (The regress)*.* One may add a second level, $\mathsf{Confirmed}_2$, requiring a confirmation of the reader's confirmation by a second reader, and so on. At every finite level the argument of Theorem 7.8 shows that the top-level confirmation is not derivable without a variable of the form rok_k at that level. The regress terminates only by the introduction of an assumption, and the question of whether to trust that assumption is the one the formal system is built to expose and not to settle.

Non-Result 7.10 (The ledger is not gamed)*.* This essay does not prove that the reader's confirmations are diligent, that the ledger is not gamed by the author, or that the checks recorded

in it were performed. The ledger records what was asserted and confirmed, and it makes the assertions and confirmations visible. It cannot make them true.

> Status of H. Unproved. Theorems 7.1 through 7.8 each show that a metric of the kind the essays reject is gameable. They support the essays' decision to demote scores and keep a ledger, and they say nothing about whether the essays should be published. Non-Result 7.10 records that the ledger itself is not shown to be sound.

8 Where the Two Effects Meet

The rebound effect and the gaming of metrics are usually treated separately, and the reasons for treating them together here are specific to the procedure. The rebound effect raises the number of briefs. The gaming of metrics determines what happens to the briefs that are selected from the larger number. The two compose, and the composition has consequences that neither alone has.

8.1 Counts grow with volume

The first essay recommends that the yield of the procedure be reported as a rate over a prespecified sample and not as a count of successes. The recommendation has a formal basis.

Theorem 8.1 (Counts and rates)**.** *Draw n briefs independently, and for each brief let α be the probability that the proxy accepts it and δ the probability that the proxy accepts it while its true quality is below a standard τ, with $0 < \delta \leq \alpha \leq 1$. Then the expected number of accepted briefs is $n\alpha$, the expected number of falsely accepted briefs is $n\delta$, and the probability that at least one false acceptance occurs is $1 - (1 - \delta)^n$, which increases to 1 as $n \to \infty$. The rates α and δ are independent of n.*

Proof. Each brief is accepted with probability α and falsely accepted with probability δ, independently of the others, so by linearity of expectation the expected counts are $n\alpha$ and $n\delta$. The probability of no false acceptance among n draws is $(1 - \delta)^n$, which is strictly decreasing in n and tends to zero since $0 < \delta \leq 1$. □

Corollary 8.2 (Jevons raises the count and leaves the rate)**.** *Under Assumption 6.1, a reduction of the generation cost by the factor λ increases the number of briefs drawn by the factor $\kappa > 1$ of Theorem 6.5. Hence the expected number of accepted briefs rises by κ, the expected number of falsely accepted briefs rises by κ, and the rates are unchanged.*

A report that the generator produced more accepted briefs after it was improved would therefore be consistent with no improvement in the acceptance rate, and a report that more false acceptances were found would be consistent with no deterioration. The count rises with volume whatever the quality does. The first essay's advice to report the fraction over a fixed sample is the only report that does not move with the volume, and Theorem 8.1 is the proof that the advice is not a matter of taste.

8.2 What a single success shows

The first essay presents one worked example and says plainly that a single success among unknown trials says nothing about the frequency of success. The statement can be sharpened.

Proposition 8.3 (A success establishes only positivity)**.** *Let $\rho \in [0,1]$ be the probability that a drawn brief is developed into an essay that a careful reader accepts. The probability that at least one of $n \geq 1$ independent draws succeeds is $1-(1-\rho)^n$, which is strictly increasing in ρ, equal to zero if and only if $\rho = 0$, and positive for every $\rho > 0$.*

Proof. The probability of no success is $(1-\rho)^n$, which is strictly decreasing in ρ on $[0,1]$ and equals one at $\rho = 0$. Hence the complement is strictly increasing, vanishes at $\rho = 0$, and is positive for $\rho > 0$. □

The proposition shows that observing a verified success would exclude exactly one value of the yield, namely zero, and would be compatible with every positive value. It would be evidence that the procedure can work, and it would carry no information about how often. This is a weaker conclusion than the observation of a success is usually taken to license, and it is also a stronger one than the worked example in the first essay can supply, since the example is described in its own record as an editorial development of inspected passages and not as a recorded response of a model. The example is therefore not a verified success of the procedure, and the proposition does not apply to it.

8.3 The best-looking brief

Combining Theorems 6.5 and 7.6, the improvement of the generator increases the number of draws, and the expected proxy value of the best-scoring draw is nondecreasing in the number of draws.

Theorem 8.4 (Cheaper generation raises the apparent best)**.** *Under Assumption 6.1, suppose the author draws n_0 briefs at the baseline price and $n_\lambda = \lceil \kappa n_0 \rceil$ at the reduced price, where $\kappa > 1$ is the factor of Theorem 6.5. If the proxy values of the briefs are independent, identically distributed, integrable and not almost surely constant, then the expected proxy value of the best-scoring brief is strictly higher at the reduced price.*

Proof. Since $\kappa > 1$ and n_0 is a positive integer, $\kappa n_0 > n_0$ and so $n_\lambda > n_0$. Theorem 7.6 gives $\mathbb{E}[M_{n_\lambda}] > \mathbb{E}[M_{n_0}]$. □

By Proposition 7.7, the rise in the expected proxy value of the best brief is compatible with no rise in the true quality of the brief selected. The author who improves the generator and then reads the best-looking of the larger number of briefs will, on average, find the proxy values impressive, and the impression will be partly or wholly an artifact of the number of draws. This is the interaction: the same improvement that increases the burden of review increases the apparent quality of what is reviewed first.

Remark 8.5. The result does not condemn the practice of selecting among many briefs. Selection among many is how the procedure finds anything. The result says that the evidence for the procedure must not be taken from the selected brief's proxy value, and it supports the first essay's insistence that the proposed study fix the draws before the judgments are made and judge each against the full sources.

> Status of H. Unproved. Theorem 8.1 and Theorem 8.4 show that counts and apparent quality rise with volume independently of the true rate. They are true whether or not the essays are published, and they bear on how the essays' claims should be read and not on whether they should appear.

9 The Decision Problem

The informal debate on which this essay builds concluded that publication is defensible with three editorial changes and not without them, and that whether to wait for a small run of the generator is a judgment about how much the author values the public record against the cost of a possible revision. The conclusion is now stated as a theorem about a small decision problem. The decision problem is deliberately qualitative, because the essay has no probabilities to offer and declines to invent them.

9.1 Options and attributes

Five options are considered. *Publish now* places both essays in the public repository in their present state. *Publish with status* places them there after the three editorial changes: a status line in the first lines of each essay, a qualification of the figure in the title of the first, and demonstrations accompanying the claims about earlier code. *Hold with hash* withholds them and records a digest of both in a tagged commit. *Hold without hash* withholds them with no record. *Publish with status and run* adds to the editorial changes a small run of the generator with its model call exercised, and it is listed for completeness although the run has not been performed and the option is not presently available.

Each option is assessed on four attributes. *Record* is the strength of the public ordering of events. *Feedback* is the degree to which others can correct or build on the work. *Misreading exposure* is the degree to which a reader could mistake a design for a result. *Revision exposure* is the degree to which a later finding could force public contradiction of what was published. The first two are better when higher and the last two when lower. Each is assessed on the three-point ordinal scale $L < M < H$.

The assessments are given in Table 2. They are not derived from anything. They are the author's qualitative judgments, recorded in the form of an ordinal table so that their consequences can be traced, and the essay claims no authority for them.

9.2 Dominance

An option x *dominates* an option y if x is at least as good as y on every attribute and strictly better on at least one, where better means higher on the first two attributes and lower on the last two. An option is *Pareto-maximal* in a set of options if no option in the set dominates it.

Theorem 9.1 (No dominant option)**.** *Among the options $\{a, b, c, d, e\}$ of Table 2 the dominance relation is exactly $b \succ a$, $c \succ e$, $d \succ a$ and $d \succ b$. The Pareto-maximal options are c and d, and they are incomparable. Among the options presently available, $\{a, b, c, e\}$, the Pareto-maximal options are b and c, and they are incomparable. No option dominates every other.*

Proof. By inspection of the ten pairs. Option b equals a on Record, Feedback and Revision and is better on Misreading (M against H), so $b \succ a$. Option c equals e on Feedback, Misreading and Revision and is better on Record (H against L), so $c \succ e$. Option d equals a on Record and Feedback and is better on Misreading (L against H) and Revision (M against H), so $d \succ a$, and equals b on Record and Feedback and is better on both of the other two, so $d \succ b$. For the remaining pairs there is an attribute on which each is strictly better than the other. Option a against c: a is better on Feedback (H against L), and c is better on Misreading. Option a against e: a is better on Record, and e is better on Misreading. Option b against c: b is better on Feedback, and c on Misreading. Option b against e: b is better on Record, and e on Misreading. Option c against d: d is better on Feedback (H against L), and c is better on Revision (L against M). Option d against e: d is better on Record, and e on Misreading. Hence the pairs not related by dominance are $(a, c), (a, e), (b, c), (b, e), (c, d), (d, e)$, which together with the four dominance pairs account for all ten. An option is Pareto-maximal if it is dominated by none: a is dominated by b and d, b by d, e by c, and c and d by none. Removing d from the set leaves b undominated and c undominated. □

The theorem formalizes the shape of the informal debate. The editorial changes improve on publishing now without cost on any attribute, which is the result $b \succ a$. The decision between publishing with the changes and holding with a hash is not settled by dominance, because publication gains Feedback and loses on Misreading and Revision exposure. The run, if it were performed, would reduce the exposure of publishing, but it would not make publishing dominate holding, because holding retains the lowest Revision exposure.

Corollary 9.2 (Scalarization is a value judgment)**.** *Encode the scale as $L = 0$, $M = 1$, $H = 2$, and consider additive utilities $U = w_R$ Record $+ w_F$ Feedback $- w_M$ Misreading $- w_X$ Revision with nonnegative weights. Then $U(b) > U(c)$ if and only if $2w_F > w_M + 2w_X$, and $U(c) > U(b)$ if and only if $2w_F < w_M + 2w_X$. The dominance relation does not choose between them.*

Proof. The two options have the same Record value H. Option b exceeds c on Feedback by two scale steps (H against L), and b is worse on Misreading by one step (M against L) and on Revision by two steps (H against L). Hence $U(b) - U(c) = 2w_F - w_M - 2w_X$, whose sign decides the comparison. Both signs occur for admissible weights, for example $(w_F, w_M, w_X) = (1, 0, 0)$ and $(0, 1, 0)$. □

9.3 Deliberation has no terminating measure

The lifecycle of Section 4.3 places the decision in the live states until it is resolved by collapse or refusal. Holding is a live state, and nothing in the rules forces it to end.

Proposition 9.3 (Unbounded deliberation)**.** *For every $N \geq 1$ there is a history of the lifecycle with at least N transitions that has not reached a terminal state.*

Proof. Take the history that proceeds from open by BIND(p) with auth(p) and TRANSFORM to transformed, and then alternates VERIFY^{-} and TRANSFORM for as many rounds as desired. Each round is two transitions, and every state visited is live. □

The cycle between transformed and vfailed is the formal image of revising a text, finding that it does not pass its checks, and revising it again. The proposition is not an observation about any person. It is an observation that a calculus which separates a failed verification from a refusal, as the lifecycle does, has no termination guarantee, and that the guarantee must come from outside: from a budget, a deadline, or a decision to refuse. The essays elsewhere insist that exhausting a budget be recorded as such and not relabelled as confidence. The same insistence applies to the decision about the essays, and the present essay is not the budget.

Non-Result 9.4 (The author decides)*.* This essay does not prove that the author will, within any bound, move the commitment to publish into a terminal state. It does not prove that the terminal state will be collapsed, and it does not prove that it will be refused.

> Status of H. Unproved. Theorem 9.1 shows that no option dominates all others and that the editorial changes dominate publishing in the present state. It leaves open the choice between publishing with the changes and holding with a hash, which Corollary 9.2 reduces to a value judgment that is not in the calculus.

10 Semi-Paranoid Risk and the Case for Analyzing Unlikely Events

Two further worries, concerning dependence on a provider and the meaning of a contribution count, are treated in the sections that follow. Both concern events that may be unlikely, and an objection arises at once to analyzing them at all. The objection is that improbable events do not merit attention, that an analysis of every conceivable misfortune would never end, and that the author, who has no estimate of any probability and has declined to invent one, has no basis on which to weigh an event whose likelihood is unknown. This section answers the objection. The answer is a decision-theoretic argument that does not depend on any estimate of likelihood, and it defines a bounded form of caution, here called semi-paranoid, that analyzes unlikely events only when a precise test is passed.

10.1 The risk register

A *risk* is a tuple $r = (E, L, \kappa, \rho)$ in which E is an event, $L \geq 0$ is the loss if E occurs and nothing has been done, $\kappa \geq 0$ is the cost of a mitigation that is carried out before E is known to occur,

and ρ with $0 \leq \rho < L$ is the loss if E occurs after the mitigation has been carried out. All three quantities are measured on a common scale of value, and none is given a number in this essay. The probability $\pi \in (0, 1]$ that the event occurs within the period of interest is unknown. A *register* is a finite set of risks.

Two acts are available for each risk. *Mitigate* (denoted M) pays κ and leaves the residual loss ρ if the event occurs. *Neglect* (denoted U) pays nothing and incurs L if the event occurs. The losses in the four cases are given in Table 3.

10.2 Three decision rules

Theorem 10.1 (Threshold)**.** *The expected loss of M is less than that of U if and only if $\pi > \pi^*$, where $\pi^* = \kappa/(L - \rho)$. The threshold π^* does not depend on π.*

Proof. The expected loss of U is πL and that of M is $\kappa + \pi\rho$. Then $\kappa + \pi\rho < \pi L$ if and only if $\kappa < \pi(L - \rho)$, and $L - \rho > 0$. □

The theorem converts a question that cannot be answered, what the probability is, into a question that can: whether the probability is above a threshold that is computed from the loss, the cost of mitigation and the residual loss. An author who does not know π may still know that π^* is small, and, if it is, that π would have to be smaller still for neglect to be the better course. The threshold analysis does not need an estimate, and for the purposes of this essay it supplies the only kind of probabilistic statement that the essay is willing to make.

The theorem is a rule in expectation, and two rules that do not use π at all are also available. The *maximin* rule chooses the act with the smaller worst-case loss. The *minimax regret* rule chooses the act with the smaller worst-case regret, where the regret of an act in a state is its loss minus the least loss achievable in that state.

Theorem 10.2 (Rules under ignorance)**.** *For a risk with $0 \leq \rho < L$:*

(*i*) *the maximin rule chooses M if and only if $\kappa + \rho < L$, that is, if and only if $\pi^* < 1$;*

(*ii*) *the minimax regret rule chooses M if and only if $2\kappa < L - \rho$, that is, if and only if $\pi^* < \frac{1}{2}$ (with indifference at equality);*

(*iii*) *the expected loss rule chooses M if and only if $\pi > \pi^*$.*

In particular, every risk mitigated by the minimax regret rule is mitigated by the maximin rule, and a risk is mitigated by the expected loss rule at every $\pi > \frac{1}{2}$ if it is mitigated by the minimax regret rule.

Proof. (i) The worst-case loss of U is L and that of M is $\kappa + \rho$. (ii) In the state in which the event does not occur the least achievable loss is 0, so the regret of U is 0 and that of M is κ. In the state in which the event occurs the least achievable loss is $\min(L, \kappa + \rho)$. If $\kappa + \rho < L$, the regret of U there is $L - \kappa - \rho$ and that of M is 0, so the maximum regret of U is $L - \kappa - \rho$ and that of M is κ, and $\kappa < L - \kappa - \rho$ if and only if $2\kappa < L - \rho$. If $\kappa + \rho \geq L$, the regret of U is 0 in both states, so U is chosen, and the inequality $2\kappa < L - \rho$ fails since $\kappa \geq L - \rho > 0$. (iii) is

Theorem 10.1. For the final claim, if $2\kappa < L - \rho$ then $\kappa < L - \rho$, so $\kappa + \rho < L$, and $\pi^* < \frac{1}{2}$ so every $\pi > \frac{1}{2}$ exceeds it. □

The three rules form a scale of caution. Maximin is the most cautious and mitigates whenever the mitigation costs less than the loss it prevents, however improbable the event. Expected loss is the least cautious, since it asks that the probability be known to be high enough. Minimax regret lies between, and it has a feature that gives the term its meaning.

Definition 10.3 (Semi-paranoid rule)**.** The *semi-paranoid rule* is the minimax regret rule: mitigate if and only if $2\kappa < L - \rho$. It is paranoid in that it mitigates whatever the probability may be. It is only semi-paranoid in that it refuses every mitigation costing as much as half the loss it would avoid.

The limit on the paranoia is what keeps the analysis finite. A fully paranoid rule would recommend any mitigation that costs less than the loss, and in the presence of many risks the total cost of such mitigations could exceed the total value being protected. The semi-paranoid rule recommends only the mitigations that are cheap relative to the avoidable loss. For a register of finitely many risks it selects a finite subset by a test that can be applied risk by risk, without estimating anything about any pair.

10.3 The filters on the register

An objection to any rule that mitigates independently of probability is that it can be exploited by invented risks. A claim that some disaster of enormous magnitude will follow from some action, together with a cheap mitigation, passes the test of Definition 10.3 for any L large enough. Three filters limit the admissible risks, and each is a condition that can be checked.

Definition 10.4 (Admissible risk)**.** A risk (E, L, κ, ρ) is *admissible* if (a) the event type E is *attested*, meaning that events of that type are known to have occurred to members of a reference class to which the author belongs, whatever their frequency; (b) the loss L is *derivable from the model*, meaning that it is bounded by a quantity computed from the author's own stake, such as the weighted size of $\mathrm{Dep}(I)$ in Section 11; and (c) $2\kappa < L - \rho$.

Condition (a) excludes events that are merely conceivable. Condition (b) excludes loss figures that are asserted and not derived, and it is this condition that blocks the exploitation described above, because the loss is a function of the structure of the author's own affairs and cannot be inflated by assertion. Condition (c) is the semi-paranoid test. The essay does not cite instances to establish (a) for the events considered in Section 11, and it records the attestation as an assumption.

Proposition 10.5 (Closure)**.** *If a risk (E, L, κ, ρ) is admissible, then so is any risk with the same E and a loss, cost and residual loss (L', κ', ρ') with $L' - \rho' \geq L - \rho$ and $\kappa' \leq \kappa$ and with (b) satisfied. Hence reducing the cost of mitigation, or enlarging the loss it avoids, can only add to the admissible risks.*

Proof. Condition (a) depends only on E. Condition (c) for the new risk follows from $2\kappa' \leq 2\kappa < L - \rho \leq L' - \rho'$. □

10.4 Long horizons

A risk that is unlikely in each period need not be unlikely over the life of a record meant to persist. Let the event have probability $h_i \in [0, 1)$ in period i, independently across periods.

Lemma 10.6 (Accumulation)**.** *The probability that the event occurs at least once in the first n periods is $1 - \prod_{i=1}^{n}(1 - h_i)$. If $h_i = h > 0$ for all i, this is $1 - (1 - h)^n$, strictly increasing in n and tending to 1, and it exceeds any $q < 1$ once $n \geq \ln(1 - q)/\ln(1 - h)$. More generally, it tends to 1 if and only if $\sum_i h_i = \infty$.*

Proof. The probability of no occurrence is the product of the probabilities of none in each period, by independence. For constant h, $(1 - h)^n$ is strictly decreasing and tends to 0, and $(1 - h)^n < 1 - q$ if and only if $n \ln(1 - h) < \ln(1 - q)$, that is, $n > \ln(1 - q)/\ln(1 - h)$ since both logarithms are negative. For the general case, $\ln(1 - h_i) \leq -h_i$, so $\prod_{i \leq n}(1 - h_i) \leq \exp(-\sum_{i \leq n} h_i) \to 0$ when the sum diverges. Conversely, if $\sum h_i < \infty$ then $h_i \to 0$, so $h_i \leq \frac{1}{2}$ for all large i, and for such i one has $-\ln(1 - h_i) \leq 2h_i$, so $\sum -\ln(1 - h_i) < \infty$ and the product converges to a positive limit. □

The lemma shows that a hazard that is small and bounded away from zero becomes near certain over a long enough horizon, and that the horizon required is computed from the hazard and not from a judgment of the horizon's length. The independence and constancy assumptions are idealizations. They are stated so that their effect can be seen. A hazard that shrinks fast enough, with a convergent sum, escapes the conclusion, and the essay has no basis for asserting that any hazard it considers does.

10.5 Mitigation must precede the event

A further reason for analyzing a risk before it materializes is specific to losses of access. A mitigation of an access risk is an action performed within a service, such as adding a recovery route or a second method of sign-in, and an action within a service requires access to it. After the event, the mitigation may be impossible. The formal statement is in Section 11, and it can be anticipated here.

Proposition 10.7 (No ex-post mitigation of access loss)**.** *Let a mitigation of a risk consist in adding a method to $\mathcal{M}(v)$ for some node v, an operation that can be performed only from within the set of nodes currently reachable. If $v \in \mathrm{Dep}(I)$ and I has been lost, then the mitigation cannot be performed, and the ex-post cost κ' is not finite.*

Proof. By Theorem 11.4, $v \notin \mathrm{Acc}(\{I\})$ after the loss. An operation on v requires v to be reachable, so it cannot be performed. □

The proposition has a consequence for the evidence on which a decision might be based. A risk whose mitigation is only possible before the event is a risk for which the observation of the event is also the observation that the mitigation has become impossible. Waiting for evidence of the event, as one might wait for evidence of any other hazard before spending on

it, forfeits the opportunity to mitigate. This asymmetry between the times at which the event can be observed and the times at which it can be mitigated is the strongest reason for analyzing unlikely events of this kind, and it does not depend on the probability.

10.6 The cost of the analysis itself

The analysis has a cost, which should be counted against its benefit. If the analysis of a risk costs κ_a, it is justified only if it can change the decision, and it can change the decision only if the threshold π^* or the admissibility test is sensitive to something the analysis discovers. The analysis of this section is an instance. It is worth the cost to the extent that a precise statement of the semi-paranoid rule and its filters lets the author dismiss an unadmissible risk without further worry, and to act on an admissible one at a stated cost. The essay does not claim that the cost has been well spent.

Non-Result 10.8 (The probabilities). This essay does not estimate the probability of any event in any register, and the results of this section do not require it to. It does not prove that any risk considered in the sections that follow is admissible. That depends on attestation, on the derivation of the loss, and on the cost of mitigation, which the author is in a position to assess and the essay is not.

> Status of H. Unproved. Theorem 10.2 defines a rule that mitigates certain unlikely risks without an estimate of their probability. The rule has no bearing on H until its inputs, which concern the author's dependencies, are supplied.

11 Dependence on a Provider

The decision problem of the preceding section treated publication as a choice among options whose consequences lie inside the repository. A further consideration was raised after the section was drafted, and it concerns the consequences of publication for the author's access to services that lie outside it. An essay in the repository is critical of a large provider of storage and identity services, and the author's accounts with that provider serve as the means of signing in to other services. The author's footprint with the provider is small. The concern is that the loss of the accounts, for whatever reason, would cut access to much more than the services the provider itself supplies. A related concern is that the repository host is also the host of a project intended as an alternative to it.

The model below makes no assumption that any provider would retaliate against criticism, and it does not allege that any has. It assumes only that termination of an account is possible, as the terms of such services generally allow, and that it may occur for reasons unrelated to the content of anything the author has written, including mistaken automated enforcement. The question is how much is lost if it occurs, and the answer depends on the structure of the author's dependencies and not on any estimate of the likelihood of the event, which this essay does not offer. The justification for analyzing an event of unknown and possibly small probability is given in Section 10.

11.1 Access structures

Definition 11.1 (Access structure)**.** An *access structure* is a finite set V of nodes, each node being an account, a service, a credential or a hosted artifact, together with, for each $v \in V$, a finite set $\mathcal{M}(v)$ of *methods*, each method being a subset of $V \setminus \{v\}$ of prerequisites. A node v is *reachable* from a set $A \subseteq V$ if some method $m \in \mathcal{M}(v)$ satisfies $m \subseteq A$. A node with $\mathcal{M}(v) = \{\emptyset\}$ is a *root*: it is held by the author directly, such as a hardware key, a password held offline, or a machine in the author's possession.

For a set $X \subseteq V$ of lost nodes, let $\mathrm{Acc}(X)$ be the least subset $A \subseteq V \setminus X$ such that $v \in A$ whenever $v \notin X$ and v is reachable from A. It is the set of nodes the author can still use. It is obtained by iterating from the empty set: roots enter first, then every node one of whose methods is satisfied by the nodes already in, and so on until no more enter.

Lemma 11.2 (Loss is monotone)**.** *If $X \subseteq X'$ then* $\mathrm{Acc}(X') \subseteq \mathrm{Acc}(X)$.

Proof. Let $A_0 = \emptyset$ and let A_{n+1} be A_n together with every $v \notin X'$ reachable from A_n, and define B_n likewise for X. By induction $A_n \subseteq B_n$: the base case is trivial, and if $v \in A_{n+1}$ then $v \notin X'$, so $v \notin X$, and v is reachable from $A_n \subseteq B_n$, hence $v \in B_{n+1}$. The least fixed points are the unions. □

11.2 Dependents

Let I be a node, for instance an identity provider's account, whose loss is under consideration.

Definition 11.3 (Dependents)**.** The set $\mathrm{Dep}(I)$ of *dependents* of I is the least set $D \subseteq V \setminus \{I\}$ such that $v \in D$ whenever every method in $\mathcal{M}(v)$ contains I or a member of D.

The definition has a vacuous case: a node with no method at all is a dependent of every I, and is never reachable. Such nodes may be ignored.

Theorem 11.4 (Loss propagates to dependents)**.** $\mathrm{Acc}(\{I\}) \cap (\{I\} \cup \mathrm{Dep}(I)) = \emptyset$*. If, in addition, the prerequisite structure is* ranked*, meaning that there is a function* $\mathrm{rk} : V \to \mathbb{N}$ *with* $\mathrm{rk}(u) < \mathrm{rk}(v)$ *for every u in every method of v, and every node is reachable when nothing is lost* ($\mathrm{Acc}(\emptyset) = V$)*, then equality holds:* $\mathrm{Acc}(\{I\}) = V \setminus (\{I\} \cup \mathrm{Dep}(I))$.

Proof. For the first claim, let $A_0 \subseteq A_1 \subseteq \cdots$ be the stages of $\mathrm{Acc}(\{I\})$ as in Lemma 11.2. We show by induction that no A_n meets $S = \{I\} \cup \mathrm{Dep}(I)$. The base case is $A_0 = \emptyset$. Suppose $A_n \cap S = \emptyset$ and $v \in A_{n+1} \setminus A_n$. Then $v \neq I$, and some method $m \in \mathcal{M}(v)$ satisfies $m \subseteq A_n$, so $m \cap S = \emptyset$. If v were in $\mathrm{Dep}(I)$ then every method of v would meet S, contradicting the existence of m. Hence $v \notin S$.

For the second claim, let $v \notin S$. Since $v \notin \mathrm{Dep}(I)$ and $\mathrm{Dep}(I)$ is the least set closed under the defining condition, some method $m' \in \mathcal{M}(v)$ has $m' \cap S = \emptyset$. We show $v \in \mathrm{Acc}(\{I\})$ by induction on $\mathrm{rk}(v)$. Every $u \in m'$ has smaller rank and lies outside S, so $u \in \mathrm{Acc}(\{I\})$ by the induction hypothesis, and $v \neq I$, so v is reachable and hence in $\mathrm{Acc}(\{I\})$. The base case, minimal rank, has $m' = \emptyset$ since every prerequisite would have smaller rank. □

Corollary 11.5 (Exposure does not depend on footprint)**.** *The set* $\mathrm{Dep}(I)$ *is determined by the family of methods* $\mathcal{M}$ *alone. It does not depend on how much the author uses* I*, on the volume of data held at* I*, or on the revenue that* I *derives from the author.*

Proof. Definition 11.3 refers only to $\mathcal{M}$ and I. □

The corollary is the formal expression of the observation that a small footprint does not imply a small exposure. A provider's interest in an account is plausibly proportional to the volume it carries, and the author's loss on termination is proportional to $|\mathrm{Dep}(I)|$ weighted by the importance of each dependent. The two quantities are determined by different parts of the situation, and an author with a small footprint at I and a single sign-in method at each of a dozen other services has a large exposure to a counterparty for whom the account is of little interest. The asymmetry should not be mistaken for a prediction. It means only that the provider's interest in avoiding termination is not a reliable protection.

11.3 Mitigation

Proposition 11.6 (Independent methods remove dependence)**.** *Let* R *be a set of roots with* $I \notin R$*, and suppose every node* v *in a set* T *has a method* $m_v \subseteq R$*. Then* $T \cap \mathrm{Dep}(I) = \emptyset$*.*

Proof. Roots have the method $\emptyset$, which meets no set, so no root is in $\mathrm{Dep}(I)$. For $v \in T$, the method $m_v \subseteq R$ meets neither $\{I\}$ nor $\mathrm{Dep}(I)$, since $I \notin R$ and $R \cap \mathrm{Dep}(I) = \emptyset$. Hence v does not satisfy the defining condition, and $v \notin \mathrm{Dep}(I)$ by minimality. □

A method contained in R is a sign-in route that does not pass through I: a credential held independently, a hardware token, a recovery route held outside I, or an address at a domain the author controls. The proposition shows that it suffices for each critical service to have one such route. Removing a method can only enlarge $\mathrm{Dep}(I)$, and adding one can only shrink it, since the defining condition quantifies over all methods.

11.4 Which dependents to protect

Section 10 supplies the test for deciding which of the dependents deserve an independent route. Let $\mathrm{val} : V \to \mathbb{R}_{\geq 0}$ assign to each node the value to the author of continued access to it, and let c_v be the cost of giving a dependent $v \in \mathrm{Dep}(I)$ a method inside R. Treat the loss of access to I as one event E. If nothing is done, the loss is $L = \mathrm{val}(I) + \sum_{v \in \mathrm{Dep}(I)} \mathrm{val}(v)$. If a set $T \subseteq \mathrm{Dep}(I)$ is given independent routes, the cost is $\kappa_T = \sum_{v \in T} c_v$ and the residual loss is $\rho_T = \mathrm{val}(I) + \sum_{v \in \mathrm{Dep}(I) \setminus T} \mathrm{val}(v)$, by Proposition 11.6.

Proposition 11.7 (What a small footprint buys)**.** *For* $T = \mathrm{Dep}(I)$ *the residual loss is* $\rho = \mathrm{val}(I)$ *and the avoidable loss is* $L - \rho = \sum_{v \in \mathrm{Dep}(I)} \mathrm{val}(v)$*. Hence the semi-paranoid rule of Definition 10.3 recommends protecting all dependents together if and only if* $2 \sum_{v \in \mathrm{Dep}(I)} c_v < \sum_{v \in \mathrm{Dep}(I)} \mathrm{val}(v)$*. Treating each dependent as its own risk* $(E, \mathrm{val}(v), c_v, 0)$*, it recommends protecting* v *if and only if* $2c_v < \mathrm{val}(v)$*.*

Proof. Direct substitution into $2\kappa < L - \rho$ and the definitions above. □

The proposition shows what the smallness of the footprint contributes and what it does not. The residual loss after protection is $\mathrm{val}(I)$, the value of the account's own services, and this is small when the footprint is small. The avoidable loss, which governs whether protection is worthwhile, is the value of the dependents and is unaffected by the footprint (Corollary 11.5). The two facts together mean that an author with a small footprint and many dependents has a risk that is cheap to remove down to the footprint and that remains costly to leave. Under the admissibility conditions of Definition 10.4, the risk is admissible when the event type is attested, the loss is derived as above from the author's own stake, and the per-dependent test is passed.

11.5 Hosting a competitor on the incumbent

The same structure describes the situation of a project hosted on the platform it is meant to replace. Let P be the repository host and let Q be a hosted artifact, a repository containing the work on an alternative. If the only copy of Q is at P, then $\mathcal{M}(Q) = \{\{P\}\}$, and $Q \in \mathrm{Dep}(P)$. The repository containing the essays is in the same position, and so is every other artifact whose only copy is hosted there.

Proposition 11.8 (Replication)**.** *Let an artifact a have $\mathcal{M}(a) = \{\{p_1\}, \ldots, \{p_r\}\}$ with $p_1, \ldots, p_r$ distinct roots standing for independently held copies. Then for any set X of lost nodes with $|X| < r$, $a \in \mathrm{Acc}(X)$.*

Proof. Since $|X| < r$ and the p_i are distinct, some $p_i \notin X$. That root is in $\mathrm{Acc}(X)$, so a is reachable by the method $\{p_i\}$. □

The version control system on which the repository is built stores the full history in every clone, so replication of the repository is inexpensive, and the proposition applies when copies are kept on independent hosts or machines. A conflict of interest between the author and the host is not assumed. The structure is the same whether or not it exists, and the remedy, replication, is the same.

Corollary 11.9 (A record outside the host)**.** *If the digest recorded in a tagged commit is also held at a node of R that is independent of the host, then the* Record *attribute of the option* Hold with hash *survives the loss of the host. If the digest exists only at the host, its Record value is conditional on the host's continued availability to the author, and on the host's willingness to attest to the ordering.*

The corollary refines the assessment of the preceding section, in which Table 2 assigned the Record attribute the value H to every option except holding without a hash. The assessment presupposed that the ledger of Lemma 4.8 remained accessible and credible. The ledger is an append-only structure only if its custodian does not rewrite it, and the guarantee for a third party is the independent copy of the digest and not the structure itself.

11.6 An additional attribute

Add to the four attributes of the decision problem a fifth, *Provider exposure*, the degree to which the option increases the chance that a provider on which the author depends takes an action against the author's access, weighted by $|\mathrm{Dep}(I)|$. The attribute is better when lower. Write $\Delta_Q \in \{0, 1, 2\}$ for the number of scale steps by which a publishing option's Provider exposure exceeds that of holding. The attribute applies to essays critical of a provider, such as an essay on metered storage that names the pricing and storage policies of a particular company, and it does not apply with equal force to the two essays that are the subject of the present decision, which discuss no provider. The extension is stated for the general case.

Proposition 11.10 (The extra attribute cannot favor publication)**.** *Adding Provider exposure to the decision problem leaves the dominance relation of Theorem 9.1 unchanged whenever the compared options have equal Provider exposure, and never causes a publishing option to dominate a holding option. Under an additive utility with nonnegative weights, the condition for preferring b to c becomes $2w_F > w_M + 2w_X + \Delta_Q\, w_Q$.*

Proof. The attribute is equal among the publishing options a, b and d, and equal among the holding options c and e, so every dominance relation within each group is unchanged. Between the groups, a holding option has Provider exposure no greater than a publishing option, and the holding options c and e were already better than the publishing options on at least Misreading exposure, so no publishing option dominates one of them. The utility difference $U(b) - U(c)$ acquires the additional term $-\Delta_Q w_Q$. □

Proposition 11.11 (Mitigation restores the original comparison)**.** *If $\Delta_Q = 0$, which holds under the conditions of Proposition 11.6 for the critical services, the comparison reduces to that of Corollary 9.2.*

The two propositions state the effect of the extra consideration and the means of removing it. The attribute shifts the balance away from publishing an essay critical of a provider unless the dependence has first been reduced, and the reduction is available independently of the essay: it consists of the additional sign-in routes and replicas of Propositions 11.6 and 11.8. The essay does not assess the cost of making them.

Non-Result 11.12 (Exposure). This essay does not prove that any provider would take or would refrain from taking action against an account, that the author's dependence on any provider is as modelled, or that the quantities $|\mathrm{Dep}(I)|$ and Δ_Q take any particular values. The access structure is the author's to specify, and the theorems say only what follows from it.

Status of H. Unproved. Theorem 11.4 and Proposition 11.10 show that dependence on a provider can enlarge the cost of publishing an essay critical of it, and that independent sign-in routes and replicas remove the enlargement. They do not show that the two essays under discussion are critical of any provider, and they do not decide H.

12 The Contribution Graph

A second consideration concerns the display of activity on the repository host. A contribution graph shows, for each day, a quantity derived from the number of commits, and it is natural to read it as a record of work. Such a graph is a metric in the sense of Section 7, and its validity depends on a convention that the graph does not itself state: what a normal number of commits per unit of work is. In some places the convention is a handful of commits per day for a day's work, and in others, it has been reported, the expected volume is far higher. The model below is indifferent to which places these are and asserts nothing about any country. It takes the existence of differing conventions as an assumption and works out what follows.

12.1 Counts relative to a convention

Let a person's genuine work per period be $w > 0$ in some unit, and let $k > 0$ be the number of commits that the person's context produces per unit of work: the granularity of commits, determined by tooling, local expectation and the practice of splitting changes. The observed count is $m = k\,w$.

Theorem 12.1 (Counts across conventions)**.** *Let two persons have granularities k_1, k_2 and work w_1, w_2. Then $m_1 > m_2$ if and only if $k_1/k_2 > w_2/w_1$. Consequently the ordering of the counts agrees with the ordering of the work for every pair (w_1, w_2) if and only if $k_1 = k_2$.*

Proof. The first claim is the inequality $k_1 w_1 > k_2 w_2$ rearranged. For the second, if $k_1 = k_2$ then $m_1 > m_2$ if and only if $w_1 > w_2$. If $k_1 > k_2$, choose $w_2 > w_1$ with $w_2/w_1 < k_1/k_2$, for example $w_2 = w_1\sqrt{k_1/k_2}$; then $m_1 > m_2$ although $w_1 < w_2$. The case $k_1 < k_2$ is symmetric. □

The theorem says that a count can be compared across persons only if the persons share a granularity, and that otherwise the comparison reflects the ratio of granularities as much as the ratio of work. The graph is a valid indicator within a population that shares a convention, and it is not a valid indicator across populations that do not. The difficulty is a failure of calibration and not of honesty: no one in a high-count context need be inflating anything relative to local norms.

Corollary 12.2 (Within-class validity)**.** *If all persons in a class share the same k, the ordering of their counts is the ordering of their work.*

12.2 Saturation of the display

Suppose the display maps counts to a small number of levels by thresholds $0 < \theta_1 < \cdots < \theta_s$, as many displays of this kind do. The assumption is made for the sake of argument, and the essay does not assert that any particular host's display has this form.

Lemma 12.3 (Indistinguishable at the top)**.** *For any class with granularity k, every period with work $w \geq \theta_s/k$ is displayed at the top level. Hence if $k \geq \theta_s/w_{\min}$, where $w_{\min}$ is the least work in an active period, all active periods are displayed identically.*

Proof. The count is $m = kw \geq \theta_s$ for $w \geq \theta_s/k$, and the top level is assigned to all counts at or above θ_s. The second claim follows by taking $w \geq w_{\min}$. □

A population with a large k therefore loses the resolution of the display, and an observer cannot distinguish a light period from a heavy one, which is the opposite of the purpose of the graph.

12.3 Splitting as a gaming channel

Granularity can be raised deliberately. A change that could be one commit can be made as many, at little cost, and an empty or trivial commit costs almost nothing. This is a gaming channel in the sense of Theorem 7.1, with a rate γ at which expenditure on splitting raises the count. When $\gamma > 1$, that is, when a unit of effort spent on splitting yields more commits than the same effort spent on work, the maximizer of the count is the allocation that spends everything on splitting. The graph therefore rewards the gaming of the graph, whenever the graph is rewarded at all.

Corollary 12.4 (The graph is extremally gameable)**.** *Let the proxy be the commit count $m = kw + \gamma b$ with target w, and let $\gamma > k$. Then the proxy is maximized under a fixed effort budget only by the allocation with $w = 0$.*

Proof. Theorem 7.1 with effort in place of the budget, with genuine rate k and gaming rate γ, after rescaling the target by the factor k. The condition $\gamma > k$ is the condition that the gaming channel is the cheaper source of count. □

12.4 Consequences for the decision

Two consequences are drawn. First, the commits that publication adds are not evidence of anything about the essays. By Theorem 8.1 and Corollary 8.2, the count of commits rises with volume independent of the rate at which the content is good, and by Theorem 12.1 no comparison of the author's count with another population's is valid unless the conventions are the same. A consideration of the form 'publishing would make the graph look better' is a consideration about a proxy, and the essay assigns it no weight.

Second, the graph has a role as a Record, in the sense of the Record attribute of Section 9, that is limited by the same convention. A graph shows that something was committed on a given day, which is the sense in which the repository is a timestamped portfolio. It does not show how much work was done or whether it was valuable, and it carries even its limited evidential force only if the history has not been rewritten. The independent copy of the digest in Corollary 11.9 supports the second condition and not the first.

Non-Result 12.5 (The graph). This essay does not prove that any host's contribution display is thresholded, that any region has any particular granularity of commits, or that the author's own count reflects anything other than the author's practice of committing work as it is produced.

Status of H. Unproved. Theorem 12.1 shows that commit counts are comparable only within a shared convention, and Corollary 12.4 shows that the count is maximized by gaming when splitting is cheap. Neither bears on whether the two essays are warranted. The graph is a proxy, and the essay assigns the proxy no weight.

13 Independence of the Hypothesis

The preceding sections have proved a sequence of conditional results. This section collects the assumptions on which they rest, states them as the axioms of a small theory, and proves the result the essay was written to be unable to avoid: that the hypothesis H is independent of those axioms, and that under them it is equivalent to a conjunction of three editorial conditions that the axioms do not decide.

13.1 The signature and the axioms

The theory is propositional. Its atoms are the following, each standing for a statement about the present state of the two essays or about the world as the preceding sections model it.

Atom	Reading
R	A reduction in generation cost raises the review burden (Theorem 6.5).
M	A proxy with a cheaper gaming channel is gameable (Theorems 7.1 and 7.8).
C	The author's review capacity is at or below the arrival rate of briefs.
K	A backlog of unreviewed briefs accumulates.
S	Each essay states its status in its first lines.
N	The title of the first essay does not state its figure without qualification.
D	The claims about earlier code are accompanied by their demonstrations.
X	The risk that a reader mistakes a design for a result is reduced.
Cr	Some essay in the publication set is critical of a provider on which the author depends (Section 11).
Z	The author's access to dependent services is secured against loss of that provider (Proposition 11.6).
H	Publishing the two essays is warranted.

The axioms are the following, labelled A1 to A5.

$$
\begin{aligned}
&\mathsf{A1}: \quad R \\
&\mathsf{A2}: \quad M \\
&\mathsf{A3}: \quad C \to K \\
&\mathsf{A4}: \quad (S \wedge N \wedge D) \to X \\
&\mathsf{A5}: \quad H \leftrightarrow \big(S \wedge N \wedge D \wedge (Cr \to Z)\big)
\end{aligned}
$$

Axioms A1 and A2 are theorems of the preceding sections, recorded as axioms because the theory is concerned with their consequences and not with their proofs. Axiom A3 is Theorem 6.8 in qualitative form. Axiom A4 is the claim that the three editorial changes lower the

exposure to misreading, which is the content of the result $b \succ a$ in Theorem 9.1. Axiom A5 is the conclusion of the informal debate stated as a biconditional, with the proviso of Section 11: publication is warranted if and only if the three editorial changes have been made and, where an essay is critical of a provider on which the author depends, the dependence has been secured. For the two essays under discussion, which discuss no provider, *Cr* is false and the proviso is vacuous; the atom is retained so that the theory covers the repository as a whole.

Remark 13.1 (The weakest link). Of the five axioms, A5 is the only one that is not a theorem of an earlier section, and it is the only one in which H occurs. It is assumed and not proved. It is the single point at which the essay's apparatus touches the hypothesis, and the touch is by stipulation. If A5 is rejected, the theory below has no axiom about H at all, and the independence result is immediate for a more trivial reason.

13.2 Equivalence and independence

Let Γ_0 denote the context declaring the eleven atoms as propositions and the five axioms as hypotheses, in the notation of Section 3.

Theorem 13.2 (Equivalence)**.** $\Gamma_0 \vdash H \leftrightarrow \big(S \wedge N \wedge D \wedge (Cr \to Z)\big)$. *In particular* $\Gamma_0 \vdash (Cr \wedge \neg Z) \to \neg H$, *and* $\Gamma_0 \vdash \neg Cr \to \big(H \leftrightarrow (S \wedge N \wedge D)\big)$.

Proof. The first claim is the axiom A5 itself. For the second, if $Cr \wedge \neg Z$ then $Cr \to Z$ is false, so the right side of A5 is false and H is false. For the third, if $\neg Cr$ then $Cr \to Z$ is true, and the right side reduces to $S \wedge N \wedge D$. □

The theorem is the result that the essay can prove about H, and it is the full extent of what can be proved. The remaining question is whether anything further follows.

Theorem 13.3 (Independence)**.** *Neither* $\Gamma_0 \vdash H$ *nor* $\Gamma_0 \vdash \neg H$.

Proof. The argument is by exhibiting two models and appealing to soundness. A *model* of Γ_0 is an assignment v of truth values to the eleven atoms under which each axiom evaluates to true. The calculus is sound with respect to the interpretation in which Prop denotes the two-element set of truth values and proofs are irrelevant, by the proof-irrelevant set-theoretic model of the calculus of constructions [8]. Hence if $\Gamma_0 \vdash \varphi$ then φ is true in every model of Γ_0.

Let v_1 assign true to all eleven atoms. Then A1 and A2 hold since R and M are true; A3 holds since K is true; A4 holds since X is true; and A5 holds since both sides are true. Let v_2 assign true to all atoms except S and H, which it assigns false. Then A1 through A3 hold as before; A4 holds since its antecedent $S \wedge N \wedge D$ is false; and A5 holds since both sides are false. Let v_3 assign true to all atoms except Z and H, which it assigns false. Then A1 through A4 hold as before, and A5 holds since $Cr \to Z$ is false in v_3, so both sides are false. The three models are summarized in Table 4.

Since H is true in v_1, $\neg H$ is false in a model of Γ_0, so $\Gamma_0 \nvdash \neg H$. Since H is false in v_2, $\Gamma_0 \nvdash H$. □

Corollary 13.4 (Relative consistency)**.** $\Gamma_0 \nvdash \bot$, *relative to the set-theoretic metatheory in which the model of the calculus is constructed.*

Proof. The model v_1 makes every axiom true and $\bot$ false. Soundness gives the claim. □

Corollary 13.5 (The atoms are free)**.** *Each of S, N, D, Cr and Z is independent of Γ_0, and so is each of C, K and X.*

Proof. For S, N and D: the model v_1 makes each true, and the model that differs from v_1 in that atom and in H (setting both false) satisfies A5, since the right side is false when any of S, N, D is false, and satisfies A4 vacuously, so each of these atoms is true in one model and false in another. For Z: the model v_3 makes it false and v_1 makes it true. For Cr: setting it false in v_1 gives a model, since $Cr \to Z$ is then true and A5 holds as before, and v_1 makes it true. For C: setting C false satisfies A3 vacuously, and C true with K true satisfies it. For K and X: setting either false is consistent with the axioms when C and $S \wedge N \wedge D$ respectively are false, and setting it true is consistent in every case. □

The corollary says that the facts on which the hypothesis turns are not facts of the theory. Whether the status lines are present is a fact about a file, whether the author's review capacity is exceeded is a fact about a person, whether the author's accounts are secured is a fact about a set of settings at services outside the repository, and whether a reader would be misled is a fact about readers. The theory has no axiom that decides any of them, and the essay has no authority to supply one.

13.3 Why the independence is not a defect

It may be thought that independence results are negative. In the present case the negative reading is wrong. A proof that H is independent of Γ_0 is a proof that H is not settled by the considerations the essay formalizes: the rebound effect, the gaming of metrics, the combinatorics of the corpus and the structure of the decision. It follows that a person who wishes to decide whether to publish cannot do so by finishing the analysis. The decision must be made by adding an axiom, which is to say by making an editorial change or declining to, and the theory records exactly which axioms (the atoms S, N and D) would suffice.

Non-Result 13.6 (Warrant). The essay does not prove H, and by Theorem 13.3 it cannot from Γ_0. It does not prove $\neg H$, and by the same theorem it cannot. It does not prove that A5 is correct, and it does not claim to.

Status of H. Unproved, and shown unprovable from the stated axioms. Theorem 13.3 proves that neither H nor $\neg H$ follows from Γ_0. Theorem 13.2 proves that H is equivalent to $S \wedge N \wedge D \wedge (Cr \to Z)$ under the stipulation A5, which is itself unproved.

14 The Self-Endorsing Sentence

A formal essay on whether to publish a formal development invites a particular temptation, and the logic of the temptation is well understood. The development could declare that whatever it proves is warranted, and then point to a proof of the hypothesis. Löb's theorem governs what happens when a theory is allowed to say this of itself, and the result is a reason for the essay's reticence [6].

14.1 Provability and the theorem

Let T be a consistent, recursively axiomatized theory that interprets enough arithmetic to define a provability predicate $\mathsf{Prov}_T(x)$, where x ranges over codes of sentences. The calculus of inductive constructions with the axioms of Section 13 is such a theory, since it interprets Heyting arithmetic, and it is recursively axiomatized by a finite list. Löb's theorem has an intuitionistically valid proof, so it applies to the theory in the form below.

Theorem 14.1 (Löb [6])**.** *For every sentence φ, if $T \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$ then $T \vdash \varphi$.*

The theorem is cited and not proved here. Its content is that a theory can establish its own soundness for a sentence only for sentences it already proves. Gödel's second incompleteness theorem is the special case $\varphi = \bot$: if $T \vdash \mathsf{Prov}_T(\ulcorner \bot \urcorner) \to \bot$, that is, if T proves its own consistency, then $T \vdash \bot$.

14.2 The temptation

The self-endorsing axiom for the hypothesis would read $\mathsf{Prov}_T(\ulcorner H \urcorner) \to H$: if this development proves H, then H. It looks modest. It does not assert H. It asserts only the soundness of the development with respect to H, which a reader might think any author of a formal development would accept.

Two facts show why the modesty is misleading. They concern a theory T^* that contains its own self-endorsing axiom, which is possible by the recursion theorem: let T^* be T together with the single axiom $\mathsf{Prov}_{T^*}(\ulcorner H \urcorner) \to H$, in which T^* refers to its own axiomatization.

Proposition 14.2 (Self-endorsement proves the hypothesis)**.** $T^* \vdash H$.

Proof. T^* has the sentence $\mathsf{Prov}_{T^*}(\ulcorner H \urcorner) \to H$ as an axiom, hence proves it. Theorem 14.1, applied to T^* in place of T, gives $T^* \vdash H$. □

Proposition 14.3 (Blanket endorsement is inconsistent)**.** *Let $T^\dagger$ be T together with the schema $\mathsf{Prov}_{T^\dagger}(\ulcorner \varphi \urcorner) \to \varphi$ for every sentence φ, with $T^\dagger$ referring to its own axiomatization. Then $T^\dagger \vdash \bot$.*

Proof. The instance with $\varphi = \bot$ is the sentence $\mathsf{Prov}_{T^\dagger}(\ulcorner \bot \urcorner) \to \bot$, which is an axiom of $T^\dagger$. Theorem 14.1 applied to $T^\dagger$ gives $T^\dagger \vdash \bot$. □

Proposition 14.2 shows that a development which endorses itself with respect to the hypothesis proves the hypothesis at once, with no argument at all. That is the cheapest possible proof of H and the reason the essay declines to give it: the proof is valid, and it establishes nothing, because all of the content has been put into the axiom. Proposition 14.3 shows that the generalization of the move to every sentence is inconsistent, so that a development cannot adopt, as a general policy, the stance that what it proves is warranted. This is the theoretical counterpart of two statements in the essays under discussion: that readiness authorizes composition and does not certify correctness, and that a model's assertion that an obligation is discharged is not a confirmation. Theorem 7.8 is the object-level form of the same restriction, in which no term of type Confirmed arises from the model's assertions alone, and Proposition 14.3 is its meta-level form.

The essay accordingly adopts no self-endorsing axiom. It does not adopt $\mathsf{Prov}(\ulcorner H \urcorner) \to H$, and it does not ask the reader to accept that the apparatus it has developed would be a reliable guide to the hypothesis if it had delivered a verdict. It has not delivered one.

14.3 The regress of deliberative essays

The two essays were followed by a debate essay about whether to publish them. This essay formalizes that one. If the question of whether to publish essay E_k is itself treated in an essay E_{k+1}, and the same question arises for E_{k+1}, the sequence $E_1, E_2, E_3, \ldots$ continues until something terminates it.

Proposition 14.4 (Regress)**.** *Let E_{k+1} be an essay on whether to publish E_k. If each E_{k+1} is written whenever E_k is held, then the sequence is infinite, and no E_k entails the termination of the sequence.*

Proof. The first claim is the hypothesis restated: the sequence continues as long as each member is held. For the second, the content of each E_k is a set of results about the publication of E_{k-1}, and none contains a result about the publication of E_{k+1} except the statement that the question arises. Termination requires a commitment in the sense of Section 4.3 to reach collapsed or refused, which is a transition of the lifecycle and not a result of any essay. □

The present essay is E_3 in this sequence, with the debate essay as E_2 and the pair of essays as E_1. The proposition does not say that E_3 should be published or held. It says that publishing E_1, E_2 or E_3 is the only act that does not require a further essay, and that holding all three is an act that does.

> Status of H. Unproved. Proposition 14.2 shows that H would follow from a self-endorsing axiom, and Proposition 14.3 shows that a general policy of self-endorsement is inconsistent. The essay adopts neither. The hypothesis remains unproved.

15 A Remark on Grounding

One consideration has been left out of the apparatus so far, because it does not fit any of the atoms of Section 13 and because the essay has no theorem to hang it on. The generation pro-

cedure that the two essays describe is not free generation. It draws a set of three documents from a finite corpus $\mathcal{K}$ of the author's earlier essays, fixes the draw by a seed, and asks for an essay assembled from those three documents and from the principles they share.

Definition 15.1 (Grounded draft)**.** Let $\tau \subseteq \mathcal{K}$ with $|\tau| = 3$ be the documents drawn for a seed s. A draft is *grounded in* τ if each of its claims is either (i) taken from a member of τ, (ii) obtained from claims of members of τ by a step that the draft states, or (iii) marked in the draft as new and unsupported.

The definition makes grounding a relation between a draft and three named documents, and it is checked by reading them. Two things follow from this, and a third does not. The first is that the review of a draft is a comparison with a bounded set of sources and not a search through everything the author has written, which lowers the review cost per draft. It does not alter the elasticity of demand, so the threshold of Theorem 6.3 is unaffected by it, and the backlog of Theorem 6.8 is changed only in its slope. The second is that the gaming channel of Section 7 narrows, since a text that reaches the output must pass through the corpus, although the entropy score of the earlier code remains as blind to word order as before. The third, which does not follow, is any conclusion about whether a grounded combination is sound, or whether its author would endorse it. A draft assembled from three true essays may assert something false of their union. Grounding certifies where the claims came from, and the obligations of the pipeline, namely accurate interpretation of the sources, an explicit statement of the connection, answers to objections and an inventory of open claims, are what say whether the assembly holds together.

It may be added, without weight, that the procedure resembles the way the corpus was produced in the first place: ideas are chosen from a large store, held together under a constraint, and written up when they cohere. The recombiner differs in making the choice explicit, repeatable from its seed and open to audit, and it does not replace the step at which a draft is read and accepted. Whether this resemblance counts toward A5 is a question the theory cannot settle, since A5 was stipulated before the consideration was raised. The consideration is a candidate reason for the stipulation. It is not a derivation of it.

Status of H. Unproved. Definition 15.1 describes a property of drafts that can be checked against their sources. It lowers the cost of review per draft and narrows one gaming channel, and it supports no theorem about H. The hypothesis remains unproved.

16 What Has and Has Not Been Established

The essay has proved nineteen results and recorded eleven non-results. The proved results establish that the count of combinations is exact under a provisional grouping rule and permits every number of essays; that addresses by integer are stable under append-only ordering and are broken by insertion; that a generator which lowers only the generation cost raises the review burden for every positive elasticity and leaves an author at capacity with a backlog that grows linearly; that the rebound in total effort vanishes at zero price if demand saturates;

that a proxy with a cheaper gaming channel is maximized only where the target is zero; that the entropy score is blind to word order and is maximized by a single repeated token; that selection on a noisy proxy overstates the target; that a model's assertions cannot yield a confirmation without a reader; that counts rise with volume while rates do not; that no option in a small decision problem dominates all others; that mitigation of an unlikely risk can be justified without an estimate of its probability, by a threshold, a minimax-regret rule and the accumulation of hazard over long horizons; that the loss of an identity provider propagates to exactly the set of dependents determined by the structure of the author's sign-in methods and not by the size of the footprint, and that independent routes and replicas remove the propagation; that commit counts are comparable only within a shared convention and are maximized by splitting when splitting is cheap; that the hypothesis is equivalent to three editorial conditions together with a proviso on dependence, under a stipulated axiom; that it is independent of the axioms; and that a self-endorsing axiom would prove it trivially while a blanket endorsement would prove everything.

None of these establishes the hypothesis. The statement can be made more exactly: each of the nineteen results holds whether the hypothesis is true or false, which is what it means for them to be proved from axioms that leave the hypothesis open. A reader who has followed the development and who still does not know whether the essays should be published has understood it correctly.

16.1 What would be needed

The quantities on which an answer would depend are empirical, and the essay lists them without estimating any. Whether the author's dependence on a provider is as modelled, and what it would cost to give each dependent an independent route, would settle whether the semi-paranoid rule recommends the mitigation. The elasticity of the author's demand for briefs, and whether it exceeds one, would settle which sign the rebound in total effort takes. The saturation level of that demand, compared with the review capacity, would settle whether the backlog grows. The ratio of review cost to generation cost would settle how large the rebound is in either case. The rate at which drawn briefs become accepted essays, measured over a prespecified sample and judged against the full sources, would settle whether the procedure the essays describe produces anything. The diligence of the reader's confirmations would settle whether the ledger discriminates. None of these has been measured, and the essay's contribution is the identification of which theorems each would feed.

16.2 On the form of the essay

The apparatus of this essay is out of proportion to its conclusion, and the disproportion is the joke. A question that the informal debate settled in a paragraph, namely that the editorial changes should precede any publication and that the rest is a value judgment, has been given a type theory, a process algebra, two economic theorems, three theorems on the gaming of metrics, a decision problem, an independence proof and a discussion of Löb's theorem, and it has come out in the same place. A formal treatment that arrives where an informal one arrived

is neither a failure nor a success. It is evidence that the informal treatment was not missing anything that the formal one could supply, and that the remaining uncertainty is not of the kind that formalization removes.

The essay's repeated disclaimer is its serious content. Each section ends with a statement that the hypothesis is unproved, and the statement is true at the end of each. The reader who wishes to know whether the two essays should be published has been given an accurate account of why this essay cannot tell them, and of the three changes that would make the question easy to answer.

STATUS OF H. Unproved. The hypothesis H is independent of the axioms of Section 13, equivalent under the stipulation A5 to $S \wedge N \wedge D \wedge (Cr \to Z)$, and not established by any result in this essay. The essay ends where it began.

References

[1] Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90.

[2] Coquand, T., and Huet, G. (1988). The calculus of constructions. *Information and Computation*, 76(2–3), 95–120.

[3] Goodhart, C. A. E. (1975). Problems of monetary management: the U.K. experience. In *Papers in Monetary Economics*, Volume I. Reserve Bank of Australia.

[4] Jevons, W. S. (1865). *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*. Macmillan.

[5] Knuth, D. E. (2011). *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley. Section 7.2.1.3, Generating all combinations.

[6] Löb, M. H. (1955). Solution of a problem of Leon Henkin. *Journal of Symbolic Logic*, 20(2), 115–118.

[7] Manheim, D., and Garrabrant, S. (2018). Categorizing variants of Goodhart's law. arXiv:1803.04585.

[8] Miquel, A., and Werner, B. (2003). The not so simple proof-irrelevant model of CC. In *Types for Proofs and Programs* (*TYPES 2002*), Lecture Notes in Computer Science 2646, 240–258. Springer.

[9] Paulin-Mohring, C. (1993). Inductive definitions in the system Coq: rules and properties. In *Typed Lambda Calculi and Applications* (*TLCA 1993*), Lecture Notes in Computer Science 664, 328–345. Springer.

[10] Strathern, M. (1997). 'Improving ratings': audit in the British university system. *European Review*, 5(3), 305–321.

A The Development in a Proof Assistant

The statements of the principal results are given below in the concrete syntax of a proof assistant in the Coq family. The listing has not been loaded into any proof assistant and has not been type-checked. It is a transcription of the statements of the essay and not a verified development, and it is presented because the form of a formal development has value as a statement of what is claimed, whether or not the development has been completed. Every proof is closed with `Admitted`, which in these systems introduces the statement as an axiom. A development in which every theorem is admitted proves nothing, and the reader is asked to take the prose proofs of the body as the substance and this appendix as its outline.

A.1 Combinatorics

```
Require Import Arith Lia.

Fixpoint choose (n k : nat) : nat :=
  match n, k with
  | _, 0 => 1
  | 0, S _ => 0
  | S n', S k' => choose n' k' + choose n' (S k')
  end.

Lemma triple_increment : forall n,
  choose (S n) 3 - choose n 3 = choose n 2.
Proof. Admitted.

Lemma group_sensitivity : forall n,
  choose (S n) 3 = choose n 3 + choose n 2.
Proof. Admitted.

Definition rank3 (c1 c2 c3 : nat) : nat :=
  choose c1 1 + choose c2 2 + choose c3 3.

Theorem rank_bijection : forall n,
  (* rank3 maps triples c1 < c2 < c3 < n bijectively onto [0, choose n 3) *)
  True.
Proof. Admitted.

Lemma append_stable : forall c1 c2 c3 n n',
  c3 < n -> c3 < n' -> rank3 c1 c2 c3 = rank3 c1 c2 c3.
Proof. Admitted.
```

A.2 The rebound effect

```
Require Import Reals.
Open Scope R_scope.

Definition demand (A eps p : R) : R := A * Rpower p (- eps).
```

```
Definition effort (A eps p : R) : R := p * demand A eps p.

Theorem jevons_threshold : forall A eps p p',
  A > 0 -> eps > 0 -> 0 < p' -> p' < p ->
  (eps > 1 -> effort A eps p' > effort A eps p) /\
  (eps = 1 -> effort A eps p' = effort A eps p) /\
  (eps < 1 -> effort A eps p' < effort A eps p).
Proof. Admitted.

Theorem review_burden_rises : forall A eps g v lam,
  A > 0 -> eps > 0 -> g > 0 -> v > 0 -> 0 < lam < 1 ->
  v * demand A eps (lam * g + v) > v * demand A eps (g + v).
Proof. Admitted.

Fixpoint backlog (Q0 a s : R) (t : nat) : R :=
  match t with
  | O => Q0
  | S t' => Rmax 0 (backlog Q0 a s t' + a - s)
  end.

Theorem backlog_unbounded : forall Q0 a s t,
  0 <= Q0 -> a > s -> backlog Q0 a s t = Q0 + INR t * (a - s).
Proof. Admitted.
```

A.3 Proxies and confirmation

```
Theorem extremal_goodhart : forall B gamma,
  B > 0 -> gamma > 1 ->
  forall a b, 0 <= a -> 0 <= b -> a + b <= B ->
    a + gamma * b <= gamma * B /\
    (a + gamma * b = gamma * B -> a = 0 /\ b = B).
Proof. Admitted.

Section Ledger.
  Variable Obl : Type.
  Variables Asserted ReaderOK Confirmed : Obl -> Prop.
  Variable model : forall o, Asserted o.
  Variable conf : forall o, Asserted o -> ReaderOK o -> Confirmed o.

  (* No closed term of type Confirmed o exists without a reader's term.
     This is a statement about proof terms; it cannot be stated inside
     the logic and is recorded here as a comment on the structure. *)
  Theorem confirmed_needs_reader : forall o,
    (forall o', ReaderOK o') -> Confirmed o.
  Proof. Admitted.
End Ledger.
```

A.4 The decision problem and the independence result

```
Inductive level := L | M | H.
```

```
Record option := { record : level; feedback : level;
                   misread : level; revision : level }.

Definition dominates (x y : option) : Prop :=
  (* at least as good on every attribute, strictly better on one *)
  True.

Theorem no_dominant_option : True.
Proof. Admitted.

Section Independence.
  Variables R M C K S N D X Cr Z Hyp : Prop.
  Hypothesis A1 : R.
  Hypothesis A2 : M.
  Hypothesis A3 : C -> K.
  Hypothesis A4 : (S /\ N /\ D) -> X.
  Hypothesis A5 : Hyp <-> (S /\ N /\ D /\ (Cr -> Z)).

  Theorem equivalence : Hyp <-> (S /\ N /\ D /\ (Cr -> Z)).
  Proof. exact A5. Qed.

  (* Independence is a statement about models, not provable here:
     neither Hyp nor ~Hyp is derivable, since both can be consistently
     added. It is recorded as Admitted. *)
  Theorem independence : True.
  Proof. Admitted.
End Independence.
```

The one proof in the listing that is not admitted is `equivalence`, which is the axiom A5 restated, and it is closed with `Qed` by `exact A5`. It is the least informative theorem of the essay and the only one that compiles.

B Catalogue of Statements Not Proved

A catalogue of the statements that the essay declines to prove is given because the reader's confusion about what has been established is the failure the essay is most concerned to prevent. The statements are those on which a reader might reasonably suppose the essay to rest. For each, the entry says why the essay does not prove it and what would have to be done to prove it.

Table 5: Statements not proved.

No.	Statement	Why not proved	What would be needed
N1	The two essays should be published.	Independent of the axioms (Theorem 13.3).	Making the three editorial changes and the proviso, then asserting A5.

Table 5: Statements not proved.

No.	Statement	Why not proved	What would be needed
N2	The two essays should not be published.	Independent of the axioms (Theorem 13.3).	Rejecting A5 and supplying a different axiom.
N3	The elasticity of demand for briefs exceeds one.	No measurement of demand exists.	Observation of the number of briefs requested at differing generation costs.
N4	Demand for briefs saturates below the review capacity.	No measurement of either quantity exists.	Observation of the author's review capacity and of the arrival rate at the lowest generation cost.
N5	The ledger is not gamed.	The ledger records assertions and confirmations and cannot record diligence.	Independent audit of a sample of confirmed obligations against the full sources.
N6	The generator yields essays at a positive rate.	A single verified success would establish positivity (Proposition 8.3), and none has been recorded.	A run with the model call exercised, judged against the full sources.
N7	The author will resolve the commitment.	The lifecycle permits unbounded deliberation (Proposition 9.3).	A decision.
N8	Some risk has a stated probability.	The essay declines to estimate any.	Not needed for the threshold analysis.
N9	A provider would act against the author's access.	No such action is assumed or alleged.	Not needed for the structural analysis, which concerns loss if the action occurs.
N10	A host's contribution display has a given form.	The essay makes no claim about any host's display.	The documentation of the host, consulted for the purpose.
N11	The axiom A5.	It is stipulated.	Argument about what warrant requires, which is outside the calculus.
N12	The development typechecks.	It has not been loaded into a proof assistant.	Loading it.
N13	The assigned ordinal values of Table 2 are correct.	They are the author's judgments.	The author's reconsideration of them.

Table 5: Statements not proved.

No.	Statement	Why not proved	What would be needed
N14	Any person or provider named or implied in the essay has behaved in any particular way.	No such claim is made.	Evidence, which the essay does not present.

C Index of Formal Results

The principal results are listed in the order of the essay, with the technique by which each is proved.

Result	Content	Technique
Lemma 3.1	Normal forms of atomic type are neutral.	Case analysis on normal forms.
Theorem 4.6	A publication requires an authorized binding, a transformation, a successful verification and no refusal.	Invariant over the transition system.
Theorem 5.2	Count of combinations.	Binomial coefficients.
Theorem 5.4	No bound relates combinations to essays.	Explicit construction.
Theorem 5.8	The combinatorial number system is a bijection.	Induction on arity with block decomposition.
Theorem 6.3	Jevons threshold.	Monotonicity of power functions.
Theorem 6.5	Review burden always rises.	Direct computation.
Theorem 6.8	An author at capacity accumulates a backlog.	Linear recurrence.
Theorem 6.10	Rebound vanishes under saturating demand.	Piecewise analysis.
Theorem 7.1	Extremal Goodhart.	Linear programming on a simplex.
Theorem 7.2	The entropy score is permutation invariant and maximized by constant texts.	Jensen's inequality.
Theorem 7.5	Selection overstates the target.	Chebyshev's association inequality.
Theorem 7.8	Self-confirmation is impossible.	Normal forms and neutral terms.
Theorem 8.1	Counts rise with volume; rates do not.	Linearity of expectation.
Theorem 9.1	No dominant option.	Exhaustive comparison.

Result	Content	Technique
Theorem 10.2	Three rules under ignorance.	Regret tables.
Lemma 10.6	Hazard accumulation.	Product bounds.
Theorem 11.4	Loss propagates to dependents.	Induction on stages of a least fixed point.
Proposition 11.6	Independent methods remove dependence.	Minimality of the dependents.
Theorem 12.1	Counts across conventions.	Rearrangement of an inequality.
Theorem 13.2	Hypothesis equivalent to editorial conditions.	Restatement of the axiom.
Theorem 13.3	Independence.	Two models and soundness.
Proposition 14.2	Self-endorsement proves the hypothesis.	Löb's theorem.
Proposition 14.3	Blanket endorsement is inconsistent.	Löb's theorem at $\bot$.

D Notation

The notation of the essay follows common practice, with the following exceptions. The symbol H denotes the hypothesis of Definition 1.1 and appears in no definition of any other symbol. The brackets $\langle o\ c_1 \cdots c_n \rangle$ denote a circle in the process algebra, and the arrow $\rightsquigarrow$ denotes a pop. The arrow $\xrightarrow{\ell}$ labelled by ℓ denotes a transition of the lifecycle of a commitment. The letters $\mathbb{E}$ and $\mathbb{P}$ denote expectation and probability, and $\binom{n}{k}$ the binomial coefficient, which is zero for $k > n$. The symbol $\rightsquigarrow$ is used for pops and not for β-reduction, which is not written separately. The word 'proof' is used in two senses, for an argument in the metatheory and for a term in the object theory, and the context disambiguates them as it does in other works of this kind. Where the two differ, the essay's results are the arguments in the metatheory.

Label	Statement	Status	Section
R1	Every addition of one work to a corpus of n works adds $\binom{n}{2}$ triples.	Proved	5
R2	The count of combinations is compatible with every number of developed essays.	Proved	5
R3	The combinatorial number system gives a bijection onto an initial segment, and an append-only ordering preserves every address.	Proved	5
R4	A reduction in generation cost always increases review burden; it increases total effort if and only if the elasticity exceeds one.	Proved	6
R5	A fixed review rate and a larger arrival rate give an unbounded backlog.	Proved	6
R6	Under saturating demand the rebound vanishes in the limit of zero price.	Proved	6
R7	A proxy with a cheaper gaming channel is maximized only at zero target value.	Proved	7
R8	Unigram entropy is permutation invariant, and the blended score is maximized by a text of one repeated token.	Proved	7
R9	Selection on a noisy proxy overstates the target.	Proved	7
R10	A closed normal term of type *Confirmed* requires a reader's confirmation.	Proved	7
R11	Absolute counts of accepted and falsely accepted briefs grow without bound in the number of draws.	Proved	8
R12	No option in the decision problem dominates every other.	Proved	9
R13	Mitigation of an unlikely risk is justified by a threshold that does not require its probability, and hazards accumulate over long horizons.	Proved	10
R14	Loss of an identity provider propagates to exactly the dependents fixed by the sign-in structure, independently of footprint, and independent routes remove it.	Proved	11
R15	Commit counts are comparable only within a shared convention and are maximized by splitting when splitting is cheap.	Proved	12
R16	H is equivalent to three editorial conditions with a proviso on dependence.	Proved	13
R17	Neither H nor $\neg H$ is derivable from the axioms.	Proved	13
R18	Self-endorsement for H proves H, and blanket self-endorsement proves a contradiction.	Proved	14
R19	A sequence of essays each deciding whether to publish the last has no internal termination.	Proved	14
N1	H.	Not proved	everywhere
N2	$\neg H$.	Not proved	everywhere
N3	The elasticity of demand for briefs exceeds one.	Not proved	6
N4	Demand for briefs saturates.	Not proved	6
N5	The ledger is not gamed.	Not proved	7
N6	The generator yields essays at any positive rate.	Not proved	5
N7	The author will resolve the commitment to publish.	Not proved	9
N8	Any risk considered has a particular probability.	Not proved	10
N9	Any provider would act against the author's access.	Not proved	11
N10	Any host's contribution display has a particular form.	Not proved	12
N11	The stipulated axiom A5.	Not proved	13

Table 1: Results and non-results. The items R1 to R19 are conditional on their stated assumptions. The items N1 to N11 are not established by this essay.

Label	Option	Record	Feedback	Misreading	Revision
a	Publish now	H	H	H	H
b	Publish with status	H	H	M	H
c	Hold with hash	H	L	L	L
d	Publish with status and run	H	H	L	M
e	Hold without hash	L	L	L	L

Table 2: Qualitative assessments of the options. Higher is better for the first two attributes and lower is better for the last two.

	Event does not occur	Event occurs
Neglect (U)	0	L
Mitigate (M)	κ	$\kappa + \rho$

Table 3: Losses for a single risk. The probability of the event is not given.

	R	M	C	K	S	N	D	X	Cr	Z	H
v_1	T	T	T	T	T	T	T	T	T	T	T
v_2	T	T	T	T	F	T	T	T	T	T	F
v_3	T	T	T	T	T	T	T	T	T	F	F

Table 4: Three models of Γ_0. The model v_1 makes the hypothesis true. The models v_2 and v_3 make it false, by a missing editorial condition and by unsecured dependence respectively.