On the Reception of the Study of Pseudo-Profound Bullshit: Definitions, Elementary Results, and an Empirical Programme for the Measurement of Unwarranted Disbelief

Flyxion
Independent Researcher

November 2026

Abstract

Pennycook and colleagues asked whether people will attribute profundity to sentences assembled at random from impressive vocabulary, and established that many will, to a degree. The present paper asks the neighbouring question of whether people will attribute vacuity to sentences whose meaning lies beyond their preparation, and establishes nothing whatever about this. It supplies definitions, several propositions whose principal virtue is that they cannot be false, an arithmetical observation concerning the original measure of discernment, explanations at three levels of preparation, and an experimental programme of considerably greater ambition than the evidence warrants, there being no evidence. Meaning is separated from the feeling of comprehension, from correctness and from evidential support, so that a verdict on one is not silently collected as a verdict on the others, and suspension of judgment is admitted as a third response beside acceptance and rejection. Further definitions concern verdicts which persist after their objects have changed and items which are dismissed before any verdict is reached. The proposed experiments vary meaning independently of accessibility, examine what instruction and provenance do to judgment, ask participants to produce the defects they claim to detect, and turn the instrument upon those who administer it. None of them has been performed. The central hypothesis is that rejection and acceptance of obscure statements can proceed from one and the same limitation. This paper does not support that hypothesis. It does not test it, and it does not establish the occurrence, prevalence or causes of anything it describes. The reader who finishes the paper believing that it has done so has made an error which the paper would be interested in studying, although it has not done so.

Keywords: incomprehension; profundity; suspension of judgment; stimulus validity; epistemic closure; research that remains to be done

1 On the Decision to Write the Original Article

It is a remarkable event in the history of inquiry when a research group composes sentences it knows to mean nothing, presents them to several hundred people with the request that each be considered carefully for its meaning, and then records with concern that meaning was found. The request deserves more attention than it has received. A participant who is handed a sentence by a university, asked to take a moment to think about what it might mean, and offered a scale on which the lowest point is a denial of any profundity at all, has been given several reasons to suppose that there is something to find. The authors themselves observe that participants may assume statements presented in a psychology study were constructed to convey something [7]. One might add that this assumption is ordinarily correct, and that a person who abandons it will have difficulty completing the consent form.

The same instruction leaves room for a participant who is simply cooperating. Asked what an odd sentence might mean, a person may reasonably take the task to be finding an interesting reading, and succeeding at it is an exercise of imagination before it is any evidence of credulity, a possibility that matters wherever aphorism, metaphor, poetry or speculation are in play. The definitions of Section 2 can accommodate this, through their treatment of context and production in the first of them. A reader who supplies a reading has in effect formed a new statement from the same string in a new context, with the reader as a second producer. The reading does not show that the generator intended it, and the indifference of the generator does not show that the reading is worthless.

None of this impugns the finding. The decision to study the matter was sound, in the way that the decision to determine whether a smoke detector responds to toast is sound. What invites comment is the confidence with which the investigation could proceed on one side of the question while leaving the other almost untouched. The meaningless stimuli were meaningless by construction. The meaningful comparison items were motivational quotations located by an internet search and selected for plain language. Discernment was then defined as the habit of preferring the second collection to the first. A participant could therefore distinguish himself by approving of sentences about persistence and rain while declining to approve of sentences about quantum abstraction, and the instrument would be unable to say whether he had detected an absence of meaning or merely an unfamiliar dialect. The authors note that the measure would be improved by a closer correspondence between the vacuous items and the meaningful controls. The programme set out below is what happens when that sentence is taken with a seriousness its authors may not have intended.

The article closes by remarking that although it “may not be truly profound, it is indeed meaningful” [7]. This claim was not administered to participants. Its profundity and meaningfulness therefore remain, by the standards of the article, matters of opinion. The same is true of the present paper, which has the additional disadvantage of not having made any observations.

2 Definitions

Obscure subject matter imposes on the investigator a duty to be obscure only on purpose. The following terms are used in the senses given here and in no others.

Definition 1 (Statement, context, reader)**.** A *statement* is a pair $x = (s, c)$ consisting of a

string s and a context c of production, the latter including who produced the string, by what procedure, and within which body of conventions. A *reader* is a pair $r = (k, d)$ consisting of a stock of competence k and a disposition d to issue verdicts.

Definition 2 (Meaning)**.** $M(x) \in \{0, 1\}$ takes the value 1 when there exist conventions, available in c and statable independently of any particular reader's impression, under which s expresses a determinate claim. A determinate claim is one for which at least some conditions of correctness or incorrectness can be given. M does not depend on r. The paper assumes that M is bivalent in order to have something to be wrong about, and acknowledges that vagueness, metaphor, and poetry are thereby treated with a brutality that a later study should repair.

Definition 3 (Felt comprehension and comprehension)**.** *Felt comprehension* $F(x, r) \in \{0, 1\}$ records whether r has the experience of understanding x. *Comprehension* $C(x, r) \in \{0, 1\}$ records whether r can derive from x the conditions of correctness that its conventions assign to it, as judged by performance on consequences, paraphrases, and counterexamples. By construction $C(x, r) = 1$ implies $M(x) = 1$. No such implication holds for F.

Definition 4 (Correctness)**.** For x with $M(x) = 1$, $V(x) \in \{0, 1\}$ records whether x is correct by the conventions of c, which for an assertion means that it is true and for an inference that it is valid. V is undefined when $M(x) = 0$. The proposition that every prime number is odd has $M = 1$ and $V = 0$, the number two being available to anyone who wishes to check.

Definition 5 (Evidentiary support)**.** For x with $M(x) = 1$ and a body of information i, $E(x, i) \in \{0, 1\}$ records whether i meets the applicable evidentiary standard for accepting x, without implying that the standard is infallible. A correct claim may lack support on the information supplied, and an incorrect claim may meet the standard on information that later proves incomplete or misleading.

Definition 6 (Certificates)**.** A *certificate* is a finite exhibit checkable by a party other than its author. A *semantic* certificate shows that $M(x) = 0$, for example that an expression is applied outside its type and that no convention of c assigns the result a reading. A *correctness* certificate shows that $V(x) = 0$ by a counterexample, a step that does not follow, or a demonstrated conflict with a conservation law. A *support* certificate shows that $E(x, i) = 0$ by identifying an obligation that i leaves undischarged. A *production* certificate shows that s was produced by a procedure that never consulted any condition of correctness. The last concerns the conduct of the producer and settles none of M, V or E.

The separation of production from product removes an apparent quarrel. A random generator of sufficient vocabulary may emit a true identity, and a scrupulous author may publish a falsehood. Pennycook and colleagues classify by the manner of production [8], while Dalton calls attention to what a sentence may do for the person who receives it [3]. These can be read as claims about two different properties of one artefact, neither of which is M.

Definition 7 (Source cue and deference)**.** A *source cue* q is information about c available to the reader, such as attribution, review history, or evidence that consequences of x have been successfully used. *Deference* is a verdict of meaningfulness resting on q without comprehension. Difficulty is not a source cue.

Definition 8 (Verdicts)**.** For each property $P \in \{M, V, E\}$ a *verdict* v_P takes one of three values: A, the property is established to hold; D, it is established to fail; and U, it is undetermined by this reader on this information. A verdict of A or D is *warranted* when supported by a demonstration or certificate that the reader is competent to assess, or by adequate source cues concerning that property; comprehension also suffices for $v_M = \mathsf{A}$. Establishing $M(x) = 0$ requires competence to inspect the expression under the relevant conventions, and cannot require comprehension in the sense defined above, there being nothing to comprehend. A warranted v_V or v_E other than U presupposes that $M(x) = 1$ has been established, whether by comprehension or by deference. *Unwarranted disbelief* is the issue of D on any property without such support.

Undefined and undetermined are different conditions. Where $M(x) = 0$, correctness and support are inapplicable, written $\perp$, which is a fact about the item and no verdict of anyone's. A conjecture that a reader cannot assess has a value of V that the reader records as U.

A reader may therefore establish that a statement is meaningful, establish that it is false, and remain undecided whether the evidence offered for it was any good. The earlier practice of recording all three under a single rejection has much to recommend it by way of brevity.

Definition 9 (Epistemic foreclosure)**.** Let competence evolve by $k_{t+1} = T(k_t, \alpha_t I_t, J_t)$, where I_t is the instruction on offer at time t, the admission variable $\alpha_t \in \{0, 1\}$ records whether the reader takes it up, J_t collects experience and learning reaching the reader by other routes, and $T(k, 0, 0) = k$. A reader is *foreclosed* with respect to a class of statements X when the reader issues unwarranted verdicts of D on members of X and those verdicts set $\alpha_t = 0$ for the instruction by which comprehension of X would be acquired.

Foreclosure closes a particular route to correction and need not halt development. Write $\pi_X(k)$ for that part of competence on which verdicts about X depend.

Proposition 1 (Persistence under refusal)**.** *Suppose the verdict on $x \in X$ depends only on x and $\pi_X(k)$, that it is* D *at k_0, that a verdict of* D *gives $\alpha_t = 0$, and that $\pi_X(T(k, 0, J_t)) = \pi_X(k)$ for every k and every t. Then $\pi_X(k_t) = \pi_X(k_0)$ and the verdict is* D *for every $t \geq 0$.*

Proof. If $\pi_X(k_t) = \pi_X(k_0)$ the verdict is D, so $\alpha_t = 0$ and $\pi_X(k_{t+1}) = \pi_X(T(k_t, 0, J_t)) = \pi_X(k_t)$. The claim follows by induction. □

The assumptions do all of the work, and the last of them is severe, since people learn things by accident. It does permit a reader to become considerably more learned in general while remaining exactly where they were on one matter, a condition for which the paper offers no examples. The case with no other routes, $J_t = 0$, satisfies it automatically. The proposition also separates foreclosure from ordinary error. A mistaken reader who still admits instruction has a route to correction, and a reader who declines further study on the strength of a certificate has made a decision about the use of time which this definition does not condemn. Foreclosure requires that an unwarranted verdict close the route by which it might have been revised.

For the lay reader. Someone who throws away every letter in an unfamiliar script, including the free lessons in how to read it, will find next year that the script is as unfamiliar as before.

In the weeds. Nothing requires I_t to be useful. Persistence follows from non-admission together with the silence of every other route on the matter at hand. The converse claim, that admission would bring revision, needs properties of T that have not been assumed and that Experiment 2 would have to investigate.

A rejection may also be correct and unwarranted at once. The physicist who dismisses a theory by exhibiting a violated conservation law and the passerby who dismisses it because the equations look unpleasant may both be right that it is false. Only one of them has done anything, and neither has shown that it means nothing.

2.1 Verdicts That Outlive Their Objects

Foreclosure begins with an unwarranted verdict. A warranted one can also close the route to its own revision. Consider a user who consulted the automated help wizards of a software system, never obtained a working answer, and sensibly stopped asking. The verdict concerned the wizards then available, on the evidence of their performance. The same user may later decline to try a quite different system that arrives under a similar description. The persistence argument above holds the matter judged fixed, and here it has been replaced.

Definition 10 (Expired verdict)**.** Let σ_t be a system at time t and x_t the claim that σ_t cannot solve problems of a stated class. A verdict affirming that claim, warranted at t_0 by information i_{t_0}, is *expired* at $t > t_0$ when it is carried over to x_t and $E(x_t, i_{t_0}) = 0$.

Proposition 2 (Silence of old evidence)**.** *Consider two histories that agree up to t_0, in one of which the system later changes its performance and in the other of which it does not. If a reader's information remains i_{t_0} in both, no verdict depending only on that information differs between them.*

Proof. The arguments of the verdict are the same in the two histories. □

An expired verdict need not be false, since a successor may be as useless as its predecessor. It has lost its support, so that the position it entitles the reader to hold about the successor is U until something further is learned. Declining to test the successor remains a possible use of limited time. Reporting the old result as a finding about the new system does not.

For the lay reader. Having learned that a shop never has what you need is a good reason to stop going. It cannot tell you what is on the shelves after the shop has changed hands.

In the weeds. Expiry is a matter of support and requires a standard for when evidence about one system transfers to another. A change of name is insufficient to defeat transfer, and a shared name is insufficient to secure it. The paper supplies no such standard and expresses no view on the merits of any particular successor.

2.2 Prior to Any Verdict

Definition 11 (Registration)**.** $G(x, r) \in \{0, 1\}$ records whether r treats x as a candidate bearer of meaning at all. The verdicts above presuppose $G(x, r) = 1$. A response issued with $G(x, r) = 0$ is recorded as N, *non-registration*, whatever words accompany it.

The first episode of the television series *Westworld* supplies an illustration. Peter Abernathy, an artificial inhabitant of a theme park, finds a photograph of a modern city and cannot let it go. His daughter Dolores, shown the same object, replies, “Doesn’t look like anything to me” [12]. The fiction presents a response installed by her makers. She has weighed no claim and found it empty, which would be D, and she holds no question open, which would be U. A reader who reports being unable to make sense of a sentence is in a different position, having registered it well enough to complain. If the response depends only on x and $\pi_X(k)$, N sets $\alpha_t = 0$, and other routes leave $\pi_X(k)$ unchanged, the induction of Proposition 1 applies with N in place of D, although no verdict has been issued and the definition of foreclosure, as written, is not satisfied. Her father registers the photograph, malfunctions, and is withdrawn from service. The ethics review for the programme below will wish to hear that no comparable outcome is anticipated.

2.3 A Precedent in Three Registers

There is a serious precedent for refusing to let one register stand for the whole of understanding. Piper Harron’s doctoral thesis on the equidistribution of lattice shapes [6], whose theorem also appears in a conventional article written with Manjul Bhargava [1], addresses the lay reader, the initiated, and the mathematician, and separates explanations for general readers from the mathematical exposition and from the detailed matter it calls weeds. It is informal and frequently very funny, and it is also a proof. This is why it is the appropriate precedent: where a playful register and a proof occupy the same document, the register cannot by itself certify a defect. One such case says nothing about how often the two keep company, or how readers respond when they do, which is the business of Experiment 6. Its prologue objects to a mathematical culture that does not encourage all ways of thinking equally. Its opening chapter compares abstraction to watching the hands of puppeteers without being able to see what the marionette is doing. A reader may therefore fail twice, by taking the hand movements to be meaningless because no marionette is in view, or by taking the tone to be unserious because it is intelligible. A bracketed editor’s note offers students a proposition that the preceding definitions cannot express: “you can successfully use math before you’ve successfully understood it” [6].

Definition 12 (Provisional use)**.** A reader makes *provisional use* of x, written $W(x, r) = 1$, when the reader correctly applies x within a stated range of tasks while $C(x, r) = 0$ and declares that this is so.

Provisional use is neither a verdict of defect nor a claim of comprehension. It permits $v_M = \mathsf{U}$ to be occupied by someone who is nonetheless getting on with the work. The declaration is essential, since its omission produces the familiar figure who operates the terminology fluently and reports having understood it.

The present paper borrows the three registers. It does not borrow the proof, having none of comparable substance to attach them to.

3 Results of an Unobjectionable Character

Proposition 3 (Impotence of the impression)**.** *Suppose there exist statements x_0, x_1 and a reader r with $M(x_0) = 0$, $M(x_1) = 1$, and $F(x_0, r) = F(x_1, r)$. Then no function of F alone*

returns M on both statements.

Proof. A function receives equal arguments and returns equal values. The values of M are unequal. □

The proposition holds equally when the shared impression is one of understanding and when it is one of bafflement, which is the whole of the symmetry this paper wishes to advertise. It does not show that any actual reader and any actual pair of statements satisfy the hypothesis. To show that would require an experiment, and the paper has been careful to avoid one.

For the lay reader. If two sentences leave you feeling exactly the same way, and one says something while the other does not, then the feeling cannot be what tells them apart. You may feel enlightened by both or defeated by both. Neither experience comes with a label.

In the weeds. The hypothesis concerns a single reader and a function of F alone. Once source cues or performance data are added to the arguments, the impossibility no longer follows from the stated premises, although nothing guarantees that the additions will resolve the matter. The experiments below add them in hope. The proof uses no property of readers, sentences, or meaning beyond the stated inequality, and would apply without alteration to the classification of melons by a person who could not see them.

Proposition 4 (Indifference of the difference score)**.** *Let Q be a reader's mean rating of the meaningful control items, B the mean rating of the vacuous items, and $S = Q - B$. A reader who assigns the minimum rating to every item and a reader who assigns the maximum rating to every item both have $S = 0$.*

Proof. In each case $Q = B$. □

The universal scoffer and the universal enthusiast are accordingly indistinguishable in sensitivity. This is a defensible property of a sensitivity measure, since neither discriminates, and it is the reason such a measure must be read alongside the levels from which it is formed. It also has consequences for correlation. For any external variable H, writing σ for a standard deviation and ρ for a correlation, and supposing σ_Q, σ_B, σ_S and σ_H to be nonzero,

$$\rho_{SH} = \frac{\sigma_Q \, \rho_{QH} - \sigma_B \, \rho_{BH}}{\sqrt{\sigma_Q^2 + \sigma_B^2 - 2\rho_{QB} \, \sigma_Q \sigma_B}}. \tag{1}$$

The identity is elementary, and the published figures obey it. In the third study, where H is performance on a battery of reasoning problems, the reported values $\sigma_Q = 0.69$, $\sigma_B = 0.90$, $\rho_{QB} = 0.38$, $\rho_{QH} = -0.12$ and $\rho_{BH} = -0.33$ yield $\rho_{SH} \approx 0.24$, against a reported 0.23, a discrepancy attributable to rounding. In the fourth study the values 0.67, 0.82, 0.43, -0.14 and -0.21 yield $\rho_{SH} \approx 0.10$, as reported [7].

The fourth case repays contemplation. Better performance on the reasoning problems accompanied lower ratings of the vacuous items and also lower ratings of the motivational quotations. The association with discernment is what remains after a larger negative association with the ratings of the vacuous items has been partly offset by a smaller one with the ratings of the controls, and in the full sample it was not statistically significant. It reached

significance in the subsample reporting no knowledge of Deepak Chopra ($\rho = 0.19$). The authors describe their support for a detection mechanism as modest, which is a word this paper would be proud to have earned. Readers preparing a replication should also know that the prose of the fourth study states the subtraction in the order opposite to that of its table and general discussion, and that its reported participant counts do not reconcile with one another. Neither observation bears on the foreclosure hypothesis, which this paper continues not to support.

For the lay reader. Suppose someone marks two exam papers and your score is how much higher they rated the good one. A marker who fails both and a marker who passes both each score nothing. Knowing the gap does not tell you whether the marker is stern or generous, and a relationship with the gap may come from either paper.

In the weeds. Equation (1) follows from $\operatorname{Cov}(Q - B, H) = \operatorname{Cov}(Q, H) - \operatorname{Cov}(B, H)$ and $\operatorname{Var}(Q - B) = \operatorname{Var} Q + \operatorname{Var} B - 2\operatorname{Cov}(Q, B)$. In particular $\operatorname{Cov}(S, B) = \operatorname{Cov}(Q, B) - \operatorname{Var} B$, so a negative association between the score and the vacuous ratings is guaranteed whenever the covariance of the two ratings falls short of the variance of the second. Its appearance in a correlation table is therefore less informative than it looks.

Proposition 5 (Confounded difficulty)**.** *Let $m_i \in \{0, 1\}$ indicate the meaningfulness of item i and $a_i \in \{0, 1\}$ its accessibility to a given population. If every item satisfies $a_i = m_i$, then in the model $y_i = \beta_0 + \beta_m m_i + \beta_a a_i + \varepsilon_i$ the coefficients β_m and β_a are not separately identifiable.*

Proof. The model depends on the two coefficients only through $(\beta_m + \beta_a)m_i$. Adding any constant to one and subtracting it from the other leaves every fitted value unchanged. □

The original stimuli do not instantiate the premise exactly, accessibility never having been measured, and the mundane and quoted items differ from the generated ones in several respects at once. The proposition is offered as a description of the limit toward which a design drifts when its sense is plain and its nonsense ornate.

For the lay reader. If every sensible sentence in a test is easy and every senseless one is hard, nobody can tell whether people are reacting to the sense or to the difficulty. The experimenter has changed two things and called them one.

In the weeds. The design matrix has two identical columns and is rank deficient. A statistical package will ordinarily respond by discarding one of them, thereby settling by column order a question that the experiment was conducted to investigate.

4 The Foreclosure Hypothesis, Which Is Not Supported Here

Four achievements should be kept apart: showing that something is possible under stated assumptions, showing that it occurs, estimating how often it occurs, and identifying what causes it. The paper concedes, since it must, that conditional demonstrations of the first kind are contributions of a sort, and that a reader who had supposed bafflement sufficient to establish vacuity may properly abandon the supposition on meeting Proposition 3. No such licence extends to the other three.

Hypothesis 1 (Common limitation)**.** *Among statements for which a reader lacks the prerequisite competence, there are readers whose verdicts on meaning shift between* A *and* D *with changes in presentation that leave M fixed, and whose rate of correct verdicts does not thereby improve.*

Hypothesis 2 (Revision under instruction)**.** *Instruction in the prerequisites of a class of meaningful statements reduces verdicts $v_M = \mathsf{D}$ on unseen members of the class more than it reduces such verdicts on matched vacuous statements.*

Hypothesis 3 (Self-obstruction)**.** *Where instruction is optional, readers who are led to issue unwarranted verdicts of* D *take it up less often than comparable readers who are not so led.*

Hypothesis 4 (Scarcity of suspension)**.** *Where the verdict* U *is offered on any property without penalty, readers use it less often than their measured comprehension would warrant.*

The motivating instances are familiar. A formalism met for the first time is dismissed as word salad, or a lay reader concludes that a clause of legal or economic jargon says nothing and exists only to impress. Neither instance has been measured here, and either dismissal may on occasion be correct.

The acceptance side of the symmetry already has an account of its mechanism. Sperber describes how trust in an authoritative source leads readers to interpret obscure remarks charitably, to take the difficulty of interpretation as evidence of depth, and, in company with other admirers, to compound the overestimate [9]. What the programme would add is a measurement of the opposite error under matched conditions: whether distrust and difficulty close off statements that instruction could open, and whether the closure persists. It adds, at present, a proposal to measure this.

A neighbouring instance concerns a verdict about production. A recalled post on a professional networking site reported that a team had spent five years building a video game, only to see it written off on release as the work of Claude Code, which its makers found deeply insulting. A commenter replied, in substance, that slop is slop whether or not humans made it. The remark rightly declines to let authorship decide quality, and then leaves quality to be found out by someone else. A low opinion of new games in general may license *declining to examine* this one, since attention is finite, with the verdict left at U. It does little to license *claiming to have judged* it. Neither remark is shown to be foreclosure in the defined sense, which requires that a verdict close a route to comprehension. As reported, the original dismissal puts an attribution of production in place of a finding about the product, and the commenter puts a general expectation in the same place. The post cannot be quoted or checked, because the perpetually refreshing feed that displayed it did not permit a second look, and whether any critic had played the game is unknown.

A third instance concerns an item that cannot be judged because it cannot be found. A low-budget film released on video around the turn of the century, in which a criminal travels back in time to rescue a young person, was remembered well enough to be described and not well enough to be named. Keyword searches of an earlier generation returned the better known films that shared its vocabulary. A later search with a fuller description produced *The Tomorrow Man* [10], in whose reported plot the rescuer is the grown and criminal self of the boy he takes, and the person who remembered the film recognized it. No verdict on meaning or merit had been involved. Shared keywords and popularity had merely stood in

for relevance somewhere between the question and the answer, and a work which is never retrieved is spared being dismissed only by being unavailable for it. The recovery shows that the obstacle could be passed in this case. It shows nothing about how often it is.

The hypothesis is worth testing because the mechanism it describes, if it occurs, would keep unfamiliar disciplines inaccessible to people who might otherwise have entered them, and that cost would merit measurement even if it were incurred rarely. This is a reason for concern and no claim of occurrence.

It must be stated plainly that nothing in the preceding sections supports any of these hypotheses. Proposition 3 shows that an impression cannot settle meaning under a stated condition, and is silent on how people behave. Equation (1) concerns the arithmetic of a score. The anecdotal observation that persons outside a discipline sometimes mock its sentences has not been collected, coded, or counted. The hypotheses are plausible, which is the property that the original research found to be so unreliable a guide, and their plausibility is accordingly entered into the record as a reason for concern.

5 Empirical Programme

5.1 Stimulus Construction

The programme begins with four families of items drawn from several disciplines: meaningful and accessible, meaningful and expertise-dependent, vacuous and accessible, and vacuous but technically dressed. Members of the second and fourth families are to be matched in length, syntax, vocabulary frequency, and density of specialist terms. A suitable meaningful item reports that the obstruction to a global section is represented by a nontrivial cohomology class, with the sheaf and the covering specified in a context withheld or supplied according to condition. Its partner exchanges the technical nouns among slots whose types forbid the exchange. Random recombination cannot be trusted to produce vacuity, since a generator with a sufficiently good vocabulary will occasionally publish a theorem. Every vacuous item therefore requires a semantic certificate in addition to a production certificate, and every meaningful item requires a statement of its conditions of correctness, a finding on whether it meets them, and a specification of the information against which its support is to be judged, each audited by two specialists who work independently. Meaningful items that are false or unsupported are included deliberately, so that rejection of a claim can be distinguished from denial that any claim was made.

A fifth family probes registration. It consists of inscriptions in novel notations that carry conditions of correctness but give the participant no familiar sign of carrying anything, each certified by showing that a separately instructed group applies those conditions reliably. Participants first describe the display freely, with no mention of meaning, claims or profundity, and only afterwards are asked for verdicts. A description which treats the inscription as ornament or damage is a candidate instance of N. The difficulty is that asking whether a thing means something presents it as a thing that might, so that the instrument is liable to destroy the phenomenon at the moment of measuring it. Adults who have agreed to take part in an experiment are, moreover, unusually disposed to suppose that what they are shown is there for a reason.

The specialists are readers. Their verdicts require warrant, which may be audited by further specialists, whose verdicts require warrant. The programme terminates this regress at

the point where the budget is exhausted and reports that point. A distinct difficulty arises at the research frontier, illustrated by the case of the two most capable readers in a field, one of whom writes something only the other could assess. If the second reader is unavailable, others need not be wholly silent, since a less prepared reader may establish a local defect and the author may hold a certificate that outsiders can check. Failing both, every honest party assigns U to the whole, and the item is excluded from the analysis for being exactly the sort of item the analysis was designed to include.

5.2 Experiment 1: Meaning Crossed with Accessibility

Participants from at least three disciplines rate items from the first four families, so that each item is expertise-dependent for some participants and routine for others. The fifth family is administered separately under the registration protocol. Each response comprises the verdict v_M, followed by v_V and v_E only where $v_M = \mathsf{A}$, a separate rating of experienced profundity, and a report of F. Comprehension C is then measured by consequence and counterexample tasks, administered after the verdicts so that the tasks do not instruct. The design separates what the original scale combines, namely whether a statement is judged to say something and whether what it says is judged to be deep. A tautology recited solemnly and a hard lemma recited flatly are expected to come apart on these two measures, although expectation is all that is available at present.

5.3 Experiment 2: Instruction and Sham Instruction

Participants first judge a set of items and are then randomly assigned either to instruction in the prerequisites of one discipline or to sham instruction of equal duration, consisting of fluent exposure to the same terms with no account of the relations among them. Both groups judge new items and rejudge a subset of old ones. The contrast of interest is the change in warranted verdicts. Sham instruction is predicted to raise felt comprehension without raising comprehension, and so to manufacture the very participant whom the original article set out to measure. That contrast concerns whether instruction can correct a verdict. It cannot show that the verdict would have kept anyone away from the instruction, since attendance was compulsory. Further participants are therefore randomly assigned to be offered the instruction, with no obligation to accept, either after issuing their verdicts or before issuing any, and uptake is recorded. Within the former group, items are presented under attributions or registers chosen to move verdicts while leaving the items fixed, so that differences in uptake between presentations do not merely reflect the greater interest or preparation of those who would have accepted in any case. This identifies an effect of presentation on uptake. To credit that effect to the verdict requires the further assumption that presentation influences uptake by no other route, which is doubtful, since a person may decline a lesson on account of its advertised source whatever was concluded about the item. An experimenter who attributes the whole difference to foreclosure will have supplied a fresh instance of verdict without certificate. An association between rejection and refusal observed without that manipulation would be reported as an association. Ethics review should be informed that one arm of the study resembles a substantial portion of ordinary professional development.

5.4 Experiment 3: Provenance

Identical items are presented with attribution to a peer-reviewed journal, to a spiritual author, to a language model, to a random generator, or to no source, with attribution crossed with meaning and difficulty. Where attributions would be false they are described to participants as hypothetical and debriefed accordingly. This addresses the possibility, raised in the original discussion, that scepticism about the source is the principal protection against unwarranted acceptance. It also addresses the less advertised possibility that the same scepticism is the principal instrument of unwarranted rejection. The authors have elsewhere maintained that bullshit is defined by its production and not by its interpretation [8], in reply to the objection that a random sentence might nonetheless strike a reader as profound [3]. A definition by production settles what the experimenter knows. Experiment 3 asks what a participant may reasonably conclude without knowing it.

5.5 Experiment 4: Produce the Defect

A participant who selects D on any property is asked to exhibit the corresponding certificate. One who selects $v_M = \mathsf{A}$ is asked to state a condition under which the item would be false or invalid, and any other affirmative verdict requires its demonstration or the source cue relied upon. Assessors blind to the original verdict score the exhibits. This procedure converts a verdict into a claim with obligations. A certificate that succeeds for a property other than the one under verdict, as when a sentence declared meaningless is shown only to be false, is credited to the property it addresses, recorded as a migration between properties, and not counted in support of the verdict given. A report of being unable to make sense of an item, offered as the ground for $v_M = \mathsf{D}$, leaves that verdict unwarranted. It is not recorded as N. It is anticipated that many participants will offer as a certificate the observation that the sentence is obviously nonsense, and that this response will be coded as a verdict wearing a certificate's coat.

5.6 Experiment 5: Reflexive Administration

Excerpts from the original article, from the present paper, and from a generated imitation of each are submitted to participants under the foregoing procedures. A finding that readers cannot distinguish this paper from its imitation would be of scientific interest and would be reported with the minimum prominence that research ethics permits. It would not show that they cannot distinguish the intellectual merits of the two, since an imitation may differ from its source along many uncontrolled dimensions, or share its merits to an embarrassing extent.

5.7 Experiment 6: Register, Tier, and Defect

A set of sound arguments and a matched set containing certified defects are each written at the three levels of preparation and in two registers, solemn and playful, giving a $2 \times 2 \times 3$ crossing of soundness, register, and tier. Participants of all three levels of preparation issue verdicts, report felt comprehension, indicate provisional use where it applies, and attempt consequence tasks. Foreclosure by register would appear as unwarranted verdicts of D on meaning or correctness that rise under playful delivery with soundness held fixed, accompanied by refusal of the offered instruction. Credulity by register would appear as unwarranted verdicts of A

that rise under solemn delivery. A further arm presents one logical problem first in a familiar social setting and then as an abstract rule about cards, following a contrast that Harron's explanation for lay readers draws [6] and that descends from the selection task [11]. It asks whether a reader who solves the first and rejects the second as gibberish revises that verdict on being shown that their structures agree. The present paper, which is playful in register and written chiefly at one tier, would be a poor stimulus and an excellent cautionary item.

5.8 Experiment 7: Sealed Paraphrase

Accessibility belongs to a text, an audience, and the instruments that audience holds, and a language model is now among the instruments. Authors write deliberately opaque passages and, before any translation is attempted, deposit a plain statement of what each is meant to say, with conditions under which it would be wrong. Matched passages with nothing behind them are produced by random recombination, each with its production certificate, and are audited for accidental theorems. All passages are previously unpublished. A model renders each into ordinary language and is also permitted to report that it recovers nothing. Judges who have not seen the translation being produced compare it with the sealed statement, and, as a control, with sealed statements belonging to other passages, since a sufficiently general paraphrase will agree with almost anything. A translation counts as *recovery* to the extent that it matches its own statement on particulars that it does not match elsewhere. Writing p_i for the ith of $n \geq 2$ sealed statements, $\widehat{p}_i$ for the translation of the corresponding passage, and K for agreement on semantic particulars fixed in advance, passage-specific recovery may be scored as

$$R_i = K(\widehat{p}_i, p_i) - \frac{1}{n-1} \sum_{j \neq i} K(\widehat{p}_i, p_j). \tag{2}$$

This is a difference score, and the programme has not forgotten Proposition 4: a translation agreeing with every statement and one agreeing with none may both receive zero, so both terms must be reported. The score is undefined for the random passages, which have no statement of their own.

Matching a sealed statement does not show that the passage supplied what was needed to reach it, since a capable model may reconstruct a likely claim from the discipline, the occasion and the familiar shape of such arguments. Two further conditions therefore give the model the same context with the passage withheld, and with the passage selectively corrupted. Recovery is credited to the text only insofar as access to the intact text improves upon both. A model that repairs a broken argument by recognising the theorem its author was probably after may do something valuable, and the value belongs to the repair.

A second group of readers, and the model in a fresh session, receive passage and translation without any statement and attempt to say whether something was recovered or supplied.

Fluency alone is expected to prove little, a reading being available for most strings to a translator disposed to provide one. Agreement with a statement sealed in advance would be considerably stronger evidence that $M = 1$. Disagreement would not show that $M = 0$, because the model may fail where a prepared human succeeds, and an author may have deposited a clearer thought than the one that reached the page. The experiment also asks what a reader holding a faithful translation now possesses, which may be comprehension, provisional use, or only deference to the translator. An author whose passage depended upon not being understood would regard a successful result as a defect in the apparatus.

5.9 Analysis

The three-valued response forbids the casual collapse of U into rejection. Correctness and support are scored only on items for which they are defined; answers elicited by mistaken affirmation of a vacuous item are retained as a record of the error on M and enter neither measure for V or E. For each property the analysis reports *coverage*, the proportion of items on which a participant commits to A or D, together with accuracy among those commitments, since a reader who is rarely wrong and almost never ventures an opinion differs from one who is somewhat less accurate about nearly everything. Suspension is neither rewarded nor penalised in itself, and instances of N are reported separately from all three verdicts. They are excluded from the detection analysis, since a reader who has not registered an item has made no decision about it, and their rate is given alongside coverage so that the exclusion remains visible. Discrimination between meaningful and vacuous items is to be estimated among committed verdicts, in the tradition of signal detection theory [5], jointly with a model of the decision to commit at all, since a reader may appear discerning by committing only on easy cases. Participants and items enter as crossed random effects, because conclusions about statements in general cannot rest on a handful of statements treated as fixed [2]. Sample sizes are to follow from a pilot, which has not been run. A result favourable to the foreclosure hypothesis would require instruction to move verdicts on meaningful items selectively. A uniform movement toward acceptance would indicate only that instruction had made participants agreeable, and the hypothesis would then have failed, as it has so far failed to do anything else.

6 Limitations

The paper reports no participants, stimuli, responses, or analyses of new data. Its definition of meaning places bivalent predicates over phenomena that are frequently graded, and its account of competence lumps every road other than the offered instruction into a single term about which nothing is known. Adequate source cues are invoked repeatedly and the threshold of adequacy is nowhere fixed. The proposed programme is expensive, and its expense may be mistaken for rigour, an error that falls outside the five families of stimuli and would require a sixth. Relevant neighbouring work already measures receptivity to nonsensical scientific statements [4], and any execution of this programme would have to build upon it and establish what the study of unwarranted rejection adds.

7 Conclusion

Sperber observed that readers are capable of judging profound what they have failed to grasp [9]. It would be strange if the same failure never produced the opposite judgment. Strangeness is not evidence, and this paper offers none. It has defined its terms, separated meaning from correctness and both from support, proved what followed from the assumptions, recomputed two published correlations from their published components, and described experiments that someone might perform. It has not shown that anybody forecloses on anything. The foreclosure hypothesis stands at the end of this investigation precisely where it stood at the beginning, except that it is now considerably better typeset.

References

[1] Bhargava, M., and Harron, P. (2016). The Equidistribution of Lattice Shapes of Rings of Integers in Cubic, Quartic, and Quintic Number Fields. *Compositio Mathematica*, 152(6), 1111–1120.

[2] Clark, H. H. (1973). The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359.

[3] Dalton, C. (2016). Bullshit for You; Transcendence for Me. A Commentary on "On the Reception and Detection of Pseudo-Profound Bullshit." *Judgment and Decision Making*, 11(1), 121–122.

[4] Evans, A., Sleegers, W., and Mlakar, Ž. (2020). Individual Differences in Receptivity to Scientific Bullshit. *Judgment and Decision Making*, 15(3), 401–412.

[5] Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

[6] Harron, P. A. (2016). *The Equidistribution of Lattice Shapes of Rings of Integers of Cubic, Quartic, and Quintic Number Fields: An Artist's Rendering*. Doctoral Dissertation, Princeton University.

[7] Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., and Fugelsang, J. A. (2015). On the Reception and Detection of Pseudo-Profound Bullshit. *Judgment and Decision Making*, 10(6), 549–563.

[8] Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., and Fugelsang, J. A. (2016). It's Still Bullshit: Reply to Dalton (2016). *Judgment and Decision Making*, 11(1), 123–125.

[9] Sperber, D. (2010). The Guru Effect. *Review of Philosophy and Psychology*, 1(4), 583–592.

[10] *The Tomorrow Man*. (2002). Directed by Doug Campbell. Video Release.

[11] Wason, P. C. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273–281.

[12] *Westworld*. (2016). Season 1, Episode 1, "The Original." Written by Lisa Joy and Jonathan Nolan. Directed by Jonathan Nolan. HBO.