

Pursuitworthiness of Breakthrough Experiments: Transforming Emerging Intuitions into Established Norms

Szymon Miłkoś

szymon.milkos@gmail.com; szymonmilkos.com

Institute of Philosophy and Sociology of Polish Academia of Sciences, Warsaw, Poland

Preprint to be published in the European Journal of Philosophy of Science.

Please cite the final version.

Abstract:

Pursuitworthiness accounts face a fundamental dilemma when confronting breakthrough experiments: investigations judged unworthy by current standards that, upon success, transform those very standards. I show how recent pursuitworthiness accounts exclude breakthroughs, undercutting scientific gain and stagnating our understanding of pursuitworthiness. The alternative, abandoning standards entirely, increases scientific waste and embraces nihilism. I demonstrate that attempts to mitigate this dilemma fail, and that its harms in scientific practice are real, severe, and compounding. I distinguish three modes of pursuitworthiness at the forefront of science: emergent (opaque intuitions), established (justified norms), and transformative (explicating emergent into established forms). I argue that transformative pursuitworthiness standards allow us to escape from the escalating dilemma into complementary advancement. While we cannot evaluate breakthrough experiments using established standards, we can evaluate how researchers explicate their opaque intuitions into justified norms. I propose a foundational transformative norm: prefer pursuits that document their explication practice, providing the evidence to advance pursuitworthiness itself. Finally, I suggest that this norm offers an independent mechanism for funding breakthrough experiments.

Keywords: Pursuitworthiness, Experiments, Science Funding, Scientific Practice, Metacognition, Breakthroughs

1. The Pursuitworthiness Dilemma and the Promise of Transformation

Pursuitworthiness accounts develop standards for evaluating which scientific investigations we should pursue, striving to understand why certain investigations merit pursuit over others (Fleisher, 2022; Laudan, 1977; Shan, 2020; Šešelja et al., 2012; Shaw, 2022). Pursuitworthiness *standards* have two components. Norms prescribe general principles ('prefer experiments testing precise predictions'). Judgments apply these norms to specific cases ('this experimental design tests precise predictions'). Standards gain normative force through instrumental rationality, justifying suitable means for given ends (if the given end is empirical accuracy, then we should follow such standards). Yet what counts as rational

pursuit transforms through scientific practice (Chang, 2012; Daston & Galison, 2021; Kitcher, 1993).

We must distinguish how we should evaluate pursuits from the description of how evaluation actually occurs in science. This pursuitworthiness *practice* operates across multiple scales. Individual researchers decide which articles to read, while laboratories select specific experiments. Field communities debate research directions at conferences, in position papers, and discussion sections. Editors make decisions via peer review, while institutions allocate resources through hiring and tenure. Funding agencies distribute grants and governments determine budgets. For clarity, I focus primarily on grant funding decisions.

Experiments constitute science's essential empirical moment, where scientists causally interact with objects of inquiry to gather crucial evidence (Chang, 2012; Fisher, 1935; Hacking, 1983; Mooij et al., 2020; Radder, 2003). Because such direct manipulation is resource-intensive, experiment selection is particularly consequential (Franklin, 1989; Galison, 1987; Steinle, 2002). Selection is challenging because forward-looking decisions occur under bounded rationality, constrained by limited cognition, incomplete information, time pressure, and finite material resources (funding, equipment, personnel) (Simon, 1955; Wimsatt, 2007). This creates a tension between *scientific gain*, the knowledge advancement returned from invested resources, and *scientific waste*, the expenditure of precious resources on investigations yielding little epistemic return.

Feyerabend (1975) showed that pursuitworthiness accounts face a fundamental dilemma amid breakthrough experiments: investigations that current standards judge unworthy yet, when successful, transform those very standards. The dilemma's first horn arises when standards are applied consistently: breakthroughs are rejected, limiting gain by excluding precisely those investigations that advance knowledge the most. This *conservative bias* impoverishes science and stagnates our understanding of pursuitworthiness, as we cannot learn new standards from breakthroughs we never pursue. The second horn beckons with anarchistic allure: abandon standards entirely, letting 'anything go'. Yet this increases scientific waste, squandering limited resources on investigations with no epistemic return, and embraces nihilism regarding the rational choice of pursuit.

This dilemma stems from the recognition problem: identifying which novel ideas will yield breakthroughs. I propose refocusing on the *process* of pursuit itself. Alongside asking 'Which ideas are worthy of pursuit?', we should ask 'Which pursuits are conducted in a worthy manner?'. If there are better and worse ways to pursue research ideas, we can select investigations based on the quality of their execution. To do so, however, we must first understand how pursuitworthiness decisions are actually made.

I distinguish three modes of pursuitworthiness at science's forefront. Most discussions focus on *established pursuitworthiness* standards, defined earlier as justified norms and judgments. However, at the frontier, researchers face what Schwartz (2008, p.1127) calls "absolute stupidity"— "we just don't know what we're doing". Nevertheless, they feel how to proceed, a phenomenon I term *emergent pursuitworthiness*. This takes the

form of opaque feelings (proto-judgments) and intuitions (proto-norms) guiding their pursuit (Collins, 2010; Polanyi, 1962). While emergent practice exists, its rationality remains debated (Kozlov, 2023). These emergent forms are the starting point for *transformative pursuitworthiness* practice: the process of explicating emergent forms into established ones. This is the work of turning private hunches into public reasons. Crucially, Schwartz suggests we can be “productively stupid,” implying there are better and worse ways to conduct this development.

I argue that transformative pursuitworthiness standards allow us to escape the dilemma into complementary advancement of scientific pursuits and our understanding of them. We bypass the conservative bias of the first horn by providing independent criteria for breakthrough attempts, evaluating how rigorously researchers explicate emergent into established forms. Simultaneously, we avoid the permissive bias of the second horn, as we do not abandon established pursuitworthiness but continue applying wherever appropriate. Ultimately, because the virtue of transformative pursuitworthiness lies in the development of established forms, this process continuously improves both scientific practice and our understanding of it.

My aim is to convince readers that developing complete transformative pursuitworthiness accounts merits philosophical attention. While not providing a complete account here, I establish four crucial points to justify this endeavor. Section 2 establishes its *plausibility* by clarifying the triadic framework and defending the rationality of emerging and transformative forms. Section 3 demonstrates its *importance* by showing how contemporary pursuitworthiness accounts inadvertently limit scientific gain and stagnate our understanding of pursuitworthiness. Section 4 proves its *urgency* by addressing objections that harms of conservative bias can be mitigated, or are negligible because most radical ideas fail anyway, revealing instead that in practice, these pains are severe and compounding. Finally, Section 5 demonstrates its *feasibility* by proposing a way forward through the foundational transformative norm: prefer pursuits that document their pursuitworthiness practice. Crucially, such epistemic accountability generates the evidence necessary to construct subsequent standards, offering a warranted path to escape the dilemma into the complementary advancement of both scientific practice and our understanding of pursuitworthiness itself.

2. A Triadic Framework of Pursuitworthiness

This section clarifies my framework by defining breakthrough experiments, contrasting the three modes of pursuitworthiness (Table 1), and illustrating their interrelationships. I define a *breakthrough experiment* as a successful investigation guided by emergent practice that current established standards deem unworthy, yet whose success grounds the transformative explication of the former into the latter. Galileo's work exemplifies this. He was guided by an emergent intuition that systematic observation reveals truth, and a feeling that telescopic observations instantiated this proto-norm. Conversely, established

Aristotelian standards, demanding logical demonstration from first principles, judged this empirical pursuit unworthy. Galileo's success, however, initiated the transformation of his approach into a new established norm.

Category/ Pursuitworthiness	EMERGING	TRANSFORMATIVE	ESTABLISHED
Form	Intuitions (proto-norms); Feelings (proto-judgments)	The process of explicating emerging forms into established ones.	Justified norms; Evidenced judgments
Function	Guide pursuits when necessary (established standards are absent or insufficient)	Develop pursuitworthiness itself	Evaluate which scientific investigations we should pursue
Practice (examples)	Feelings of confidence, solvability, feasibility, fruitfulness; Emotions of awe, worry; Aesthetics of beauty; Unarticulated hunches or even unconscious instincts	Pattern recognition, dialogue and debate, trial and error, scenario planning, codifying best practices, reflexive journaling, experimentation, explicit reasoning	Applying formal criteria, explicit justifications, following codified protocols
Standard (examples)	Prefer what feels fruitful based on perceived modularity of experimental design promising easy adoption	Prefer pursuits that document the explication practice; Prefer networks of intertwined pursuits over isolated ones	Prefer experiments testing precise predictions; Prefer designs maximizing statistical power.
Rationality	Following emergent intuitions as half-baked hypotheses to gain feedback from mismatches between predicted and actual feelings.	Justifying how explicated standards better achieve given ends	Justified suitable means for given ends

Table 1. Differences between emerging, transformative, and established pursuitworthiness across several key dimensions, collecting distinctions that are developed throughout the paper.

To clarify the interplay within this triadic framework, I adapt the Stacey Matrix (Stacey, 1996, 2007; Zimmerman, 2001). Stacey, working in the fields of management and leadership, aimed to help organizations “generate new strategic directions” (Stacey, 1992, p. xi) by recognizing that decision-making approaches must vary with context. He mapped these situations along two dimensions: *agreement* among stakeholders regarding goals and

strategies, and *certainty* regarding cause-and-effect relationships—that is, the ability to predict intervention outcomes. Transposed to pursuitworthiness, 'agreement' becomes the *consensus* on general norms, while 'certainty' becomes the *confidence* in specific judgments.

These dimensions define a pursuitworthiness landscape with two poles (see Figure 1). In the routine zone (high norm consensus, high judgment confidence), established pursuitworthiness operates smoothly through explicit justification. Opposite lies the chaotic zone (low agreement, low certainty), which relies solely on opaque feelings and intuitions of emergent pursuitworthiness. Real-world pursuits traverse this landscape. Consider CRISPR gene editing (Doudna & Sternberg, 2017; Lander, 2016). Zhang and Doudna worked in the routine zone when applying established molecular biology norms to known systems. Extending these norms to novel applications shifted them into a region of high consensus but low confidence; standards were clear, but application was not. Conversely, the debate over germline editing represented low consensus despite high confidence, since conflicting norms clashed even as technical capabilities were understood. In fact, the entire pursuit originated in the chaotic zone, driven by inchoate feelings about bacterial immune potential.

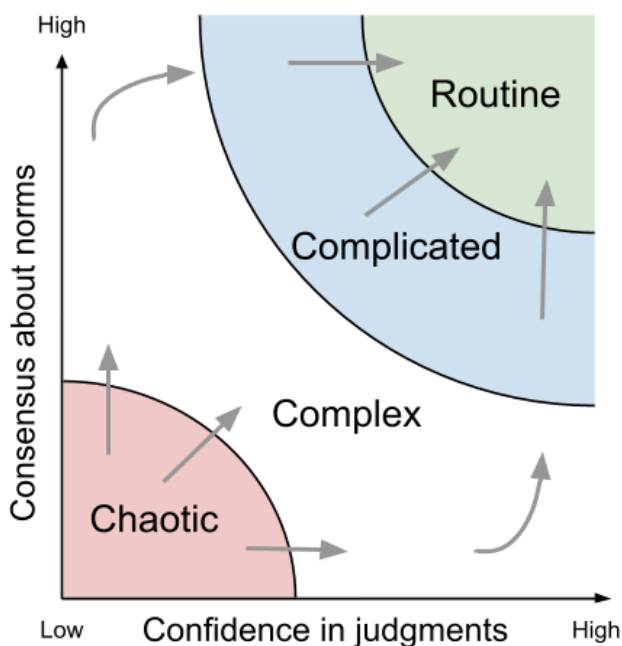


Fig 1. The Landscape of Pursuitworthiness Practice. This diagram maps the context of pursuitworthiness practice along two dimensions: consensus regarding norms and confidence in judgments. The grey arrows represent transformative pursuitworthiness practice: the directed process of explicating emergent into established forms. For zone descriptions, see text.

To guide decision-making between these poles, Stacey defined two intermediate zones. The complicated zone exists where both dimensions are at least moderate, but not simultaneously high. If consensus is lower, Stacey recommends pursuing parallel directions as a compromise; if confidence is lower—meaning we struggle to apply standards—we should focus on a clear shared vision (the given end) to help correct actions along the way. The complex zone arises when at least one dimension is low, and their sum falls below half. Here, the low dimension risks a critical flaw, and the insufficient aggregation means that any decision is no better justified than a random pick. In such cases, one should prioritize practices that increase values along these axes before deciding a direction. Hereafter, I generalize this idea.

Transformative pursuitworthiness is a directed diagonal trajectory through this matrix, moving from the chaotic to the routine zone by increasing confidence (articulating feelings into judgments) and consensus (crystallizing intuitions into norms). Each region of the landscape demands distinct explication practices: the chaotic zone requires pattern recognition to identify recurring feelings; conflicting intuitions necessitate dialogue and debate; emerging norms demand tests via trial and error; novel cases require scenario planning; and finally, approaching the routine zone, we codify best practices and explicitly justify standards.

Do emergent or transformative pursuitworthiness standards exist alongside these practices? As a preface to the positive answer I present below, consider how Fleck (1935/1981, p. 92) described the phenomenology of successful transformation by increasing “thought constraint” toward “directly perceived” facts. Asserting that scientists will “recognize an analogy” in their practice, Fleck describes the starting point as a “chaos of feeling,” where a scientist’s “feeling, will, and intellect all function together as an indivisible unit.” Here, ideas appear as an “artificial effect” of personal activity, constantly “melting away” under test. The scientist must distinguish what obeys their will from what “arises spontaneously and opposes it.” This requires seeking “resistance”—the passive element grounded in the object of inquiry—and “thought constraint”, the invariance between active cognition and that passive resistance. Explicating this constraint for others is a lengthy process of fact origination: “At first there is a signal of resistance in the chaotic initial thinking, then a definite thought constraint, and finally a form to be directly perceived” (p. 92).

The key insight for the following argument is that successful breakthroughs imply a nascent rationality within emergent pursuitworthiness. Opaque intuitions must contain novel rational content—Fleck’s ‘passive resistance’—that established standards miss. It is precisely this content that transformative pursuitworthiness explicates into new established standards.

2.1 The Rationality of Emerging and Transformative Pursuitworthiness

Rationality grounds our ability to endorse pursuitworthiness decisions as good or condemn them as bad. Traditional accounts locate such normative force in *a priori* rationality: adherence to coherent axiomatic systems like logic or probability theory. In a consequentialist approach, Schurz and Hertwig (2019) define rationality by the *cognitive success* of the system generating the decision. They characterize it as the product of ecological validity (accuracy in target environments) and applicability (scope of those environments). Instead of judging conformity to a single standard, this framework compares predictions from competing cognitive systems models. As a result, conflict between observed success (practice) and modeled predictions (standard) acts as a vital feedback signal. Assuming the task constraints (e.g., goal, resources, environment) are

correctly modeled, this implies two possibilities: either the practice is superior (requiring a better model) or the model is superior (offering rational improvements to practice).

While Schurz and Hertwig focus on externally accessible rationality, I suggest emergent pursuitworthiness reflects an internal consequentialist rationality. For the epistemic agent experiencing them, intuitions and feelings operate as distinct *internal* cognitive (sub)systems. Below, I demonstrate how the fallibility of these systems provides the agent with the feedback necessary for rational revision. This will ground the premise of nascent rationality awaiting explication.

I propose modeling emerging pursuitworthiness in terms of metacognitive feelings. These representations monitor cognitive processes and outcomes, guiding complex problem-solving despite remaining opaque (Ackerman & Thompson, 2017; Arango-Muñoz, 2014; Proust, 2013; Shea, 2023). In experimental contexts, scientists experience feelings of solvability ('this experiment will crack the problem') or feasibility ('we can pull this off with available resources'). Cognitive science shows metacognitive feelings reliably track difficulty and progress even when reasoners cannot articulate why, though they also mislead (Ackerman & Thompson, 2017). Closely related are epistemic emotions (Nerantzaki et al., 2021; Vogl et al., 2021): worry prompts assumption re-examination (Morton, 2009), while awe at phenomena exceeding current frameworks motivates breakthrough investigations rather than incremental puzzle-solving (Gottlieb et al., 2018; Valdesolo & Graham, 2014). Similarly, aesthetic feelings, such as the beauty perceived in an elegant theory, can serve as a powerful heuristic, synthesizing a scientist's sense of simplicity, explanatory fit, and unifying power into a single, motivating intuition (Miyata-Sturm, 2024). But how do they work?

Metacognitive feelings indicate emerging pursuitworthiness via three components (Arango-Muñoz, 2014; Koriati et al., 2008; Proust, 2013; Shea et al., 2014). First, vehicle-based cues: cognitive features carrying epistemic information; second, inferential appraisal: evaluating these cues against implicit standards; third, a heuristic function guiding resource allocation under bounded rationality. Shea et al.'s (2014) visual perception example illuminates this: neuronal firing rate variance (vehicle-based cue) signals reliability (inferential appraisal), with high triggering moving to the next task, and low prompting repetition or modification (heuristic), with the low/high threshold depending on importance and resources. Crucially, assessing rationality requires modeling the specific cognitive system and task constraints of a given emergent practice. Consider potential experimental examples: feasibility may stem from the processing fluency of previous similar experiments, while fruitfulness may arise from perceived modularity promising easy adaptation of an experimental design.

Scientists' emerging pursuitworthiness practice extends beyond basic metacognitive feelings, incorporating theoretical and empirical knowledge. Consider experimental confidence: vehicle-based cues include statistical measures (p-values, effect sizes); inferential appraisal integrates theoretical reasoning about validity; and the heuristic function depends critically on given ends. Ideally, epistemic pursuitworthiness maximizes scientific gain and

minimizes scientific waste. However, non-epistemic factors often dominate: publication pressure for personal career advancement acts as a structural conflict of interest (Acem et al., 2025) that drives the pursuit of easily publishable low p-values over harder experiments with larger effect sizes, sacrificing epistemic returns (Smaldino & McElreath, 2016). This tension between epistemic ideals and non-epistemic pressures is a recurring theme we will revisit.

Is it rational for scientists to follow emergent pursuitworthiness? Appeals to evolutionary reliability (Gigerenzer, 2007; Mercier & Sperber, 2017) do not extend straightforwardly to complex scientific feelings. Within a cognitive success framework, emergent forms appear more applicable than established ones, though less ecologically valid. Definitive comparison requires more data on practice, a gap our proposed transformative norm addresses (Section 5). Compounding this challenge, the cognitive task model itself may be flawed. For instance, the history of plate tectonics shows that scientists successfully prioritized the aesthetic coherence of a theory over its empirical adequacy to evidence that was later proven misleading (Miyata-Sturm, 2024). Finally, lacking clear standards, one relies on the no-alternative argument: emergent intuitions are often the sole guide, provided they outperform random choice.

I argue that following emergent pursuitworthiness is rational even if its initial guidance is no better than random—provided it is done in a particular way. My position aligns with Currie (2023), who argues the link between aesthetics and truth is not given, but forged through practices that “attune aesthetic judgment to epistemic purposes.” However, while Currie focuses on how novices learn the present attunement of a community of practice (embracing truth as a shared vision that feels good and learning what ‘elegant’ means in a given field), I focus on how this attunement originates and develops.

I suggest that emergent pursuitworthiness is internally rational because mismatches between predicted and actual feelings provide vital feedback. Drawing on manipulationist accounts of causation (Pearl, 2009; Spirtes et al., 2001; Woodward, 2003), I view experimental results as common effects of two causes: our activities (experimental design) and objects of inquiry. Similarly, metacognitive feelings have dual causes: our cognition and its objects. Peak emotional intensity occurs upon learning results (Kozlov 2023), precisely where intervention meets nature. Crucially, knowing common effect reveals information about their causes. As Kozlov (2023) shows, surprising results prompt reexamination of both experimental design (‘did we err?’) and theoretical understanding (‘were we wrong?’). Similarly, assuming no interference, conflict between predicted and observed feelings provides dual feedback. First, it allows us to develop the metacognitive components themselves: calibrating cues, validating appraisals, or adjusting heuristic functions by manipulating them to see what works. Second, analogous to theoretical anomalies, the conflict may arise from incorrect assumptions about the object, where it behaves in a way we cannot explain. Scientists must then manipulate assumptions about the object and observe which revisions withstand scrutiny. This feedback constitutes the *nascent*

rationality that is explicated through transformative practice. Therefore, scientists should follow emergent pursuitworthiness, but only when treating hunches as proto-hypotheses (to be falsified) and the processes generating these feelings as half-baked experimental designs (to be made transparent and epistemic). This marks the inception of transformative pursuitworthiness.

The rationality of transformative pursuitworthiness lies in justifying how proposed standards better achieve given ends. Consider the transformation from feelings of easiness to norms of simplicity. Scientists operating under bounded rationality adopt efficiency as a given end, initially leading to the emerging standard ‘prefer what feels easier’ when selecting experimental approaches. First explication would be just stating ‘prefer easy methods because of efficiency’ which only adds to established standards. Then, through repeated practice, patterns emerge: approaches that feel ‘easier’ correlate with structural features (fewer parameters, symmetric relations, recursive definitions), though scientists cannot initially articulate this connection. The transformative step occurs when scientists recognize that evaluating these structural features (simplicity) *before* implementation serves efficiency better than discovering ease through trial-and-error. This rational explication transforms the established standards: that ‘prefer simplicity’ overrides ‘prefer what feels easier’ by demonstrating its superior suitability for the given end of efficiency. This pattern extends across science. Physicists’ aesthetic feelings, such as Einstein’s ‘deep conviction’ about unified fields or Dirac’s insistence on mathematical beauty, transformed into symmetry principles that proved instrumentally valuable through Noether’s theorem (Brading & Castellani, 2003; McAllister, 1999). Contemporary algorithmic information theory continues this trajectory, transforming simplicity into measurable compression ratios—quantitative standards arising from initially opaque feelings (Grünwald, 2007; Li & Vitányi, 2019). Note that this rationality assumes that ‘proposed standards’ and ‘given ends’ are provided by discovery and justification accounts, respectively.

Experimental methodology shows successful transformative trajectories amid non-epistemic influences. Fisher’s (1935) randomization exemplifies the full arc: from opaque feelings (‘this comparison seems fair’) to partial articulation (‘randomization ensures fairness’) and finally established norms (‘random assignment with specific distributions’). Similar transformations produced double-blind protocols (explicating bias intuitions) (Kaptchuk, 1998) and preregistration (formalizing concerns about researcher degrees of freedom) (Nosek et al., 2018). Consider, however, the explication of certainty into statistical measures. While establishing measures of confidence was a rational step, the adoption of specific thresholds (like $p < 0.05$) and their naive application are largely shaped by non-epistemic influences. Current efforts to rethink these arbitrary thresholds and practices around them represent the next phase of explication (Amrhein & Greenland, 2018; Gigerenzer, 2004). This illustrates that non-epistemic influences do not undermine the rational trajectory of transformative pursuitworthiness.

The normative justification for transformative pursuitworthiness accounts mirrors the cognitive success account (Schurz & Hertwig, 2019). They start with a descriptive

premise, that “cognitive success is instrumental for all—or at least most—purposes”, and bridge the is-ought gap with a normative premise: “it is by-and-large good to help people attain their personal goals.” Similarly, I begin with the descriptive premise that achieving goals more efficiently is instrumental for any given end, under bounded rationality. To establish its normativity, we need to assume that it is generally good to help science achieve its goals more efficiently. This elevates efficiency (latent in the cognitive success account as ‘cognitive cost’) to a central principle. In fact, efficiency enhances effectiveness, by enabling the attainment of more goals. The key distinction is the context of bounded resources. While they focus on single tasks, I emphasize inter-task relations. This has two implications. First, selecting lower immediate returns may be rational if it promises to develop efficiency in the long term. Second, given a common finite resource pool, pursuing one experiment entails the opportunity cost of not pursuing others.

3. Partial Accounts and Differential Empowerment

Contemporary pursuitworthiness accounts remain vulnerable to the dilemma. By addressing only some respects of what makes research pursuitworthy, they create harm through *differential empowerment*: systematically empowering specific approaches leaves unsupported ones disadvantaged, especially within zero-sum funding contexts. I analyze these accounts not to dismiss their contributions, but to show how partial frameworks (3.1) decrease scientific gain and (3.2) stagnate our understanding of pursuitworthiness. While correct with targeted aspects, this partial validity drives differential empowerment, turning theoretical limitations into systemic bias.

3.1. Pains of Criticizability

DiMarco and Khalifa’s (2022) apokritic model exemplifies accounts operating exclusively through established pursuitworthiness. Evaluating research via criticizability, apokritic obligations prescribe pursuit (“if question Q about x has feature F, then scientists with capability C should pursue Q”), while prohibitions forbid it (“if Q about x has bug B, then no one should pursue Q”). This enables precise evaluation wherever features and bugs have been explicated into recognized norms. The authors frame these as standard limitations: norms are pro tanto, evaluation is contextual, and question origination is assumed. However, apokritic model drives differential empowerment: it empowers only research whose pursuitworthiness is already established, leaving other pursuits relatively disadvantaged.

This manifests in the model’s own paradigmatic case. When patients with a deadly, contagious disease respond to antibiotics, established pursuitworthiness mandates bacteriologists to ask ‘Which bacterial strain causes this disease?’ (epistemic coherence plus moral urgency), while prohibiting ‘Which virus causes this disease?’ (since antibiotic response contradicts viral causation). This exemplifies routine zone’s clarity: explicit norms

yield specific judgments. Yet breakthrough attempts often violate such prohibitions. Researchers following hunches might investigate viral involvement, potentially discovering that a virus compromised immunity, enabling secondary bacterial infection. Empowering routine investigations renders breakthrough attempts less likely, decreasing potential scientific gain.

The apokritic model systematically excludes emerging pursuitworthiness by demanding what cannot yet exist: the explicit articulation of features and bugs. While clear cases exist, many breakthrough experiments operate precisely where such clarity is absent. Researchers experiencing feelings-as-not-yet-judgments cannot identify features or bugs in their experiments, and intuitions-as-not-yet-norms resist articulation. One might object that this conflates transparency with existence. But this misses the point: these opaque forms contain genuine nascent rationality (hence breakthroughs' eventual success). This model supports only established forms where explication has already occurred.

Clearly, assuming question stability excludes breakthrough investigations. Authors maintain that an answer already is true or false at the time of pursuit—what is lacking is evidence. Yet at the pursuit frontier, questions are often ill-formed or target the wrong object, making answers 'not even wrong' (Scheel, 2022). Pursuit continuously refines its object as researchers encounter unexpected results and develop new distinctions; each experiment can transform entire networks of questions (Firestein, 2012; Gruber & Wallace, 2001). The apokritic model systematically disadvantages breakthrough attempts that challenge the very question being asked.

Finally, consider the capability-question fit requirement. In the viral-bacterial case, bacteriologists possess the 'correct' capabilities, which may blind them to non-bacterial causes, while virologists bring crucial outsider perspectives but lack the required credentials. Defenders might argue that 'other' capabilities should be employed for complex questions, but this merely pushes the problem back: how do we justify which 'other' without established reasons, relying only on opaque feelings?

3.2. How Coherence Can Lead to Stagnation

Duerr and Fischer's (2025) virtue-economic account advances beyond pure established pursuitworthiness, yet creates new forms of differential empowerment. Built on Peirce's economy of research, their framework adopts a single meta-norm: cost-benefit virtuous pursuit is worthy. Benefits derive from an idea's instantiation of virtues (e.g., empirical adequacy, explanatory power, fertility), while costs reflect vicious cognitive demands (complexity, unfamiliarity, incoherence). This framework naturally extends to experimental designs. Consider quantum entanglement experiments probing consciousness theories: they score high on fertility (opening new empirical domains) and explanatory power (potentially resolving long-standing debates) but low on familiarity (departing from established neuroscientific methods) and coherence (connecting poorly understood quantum and neural processes). The ideas' assessment and comparison are made by the ideal evaluator through deliberative judgment (emerging pursuitworthiness) grounded in a

‘triple reticulated’ (rational explication) weighted virtue matrix (established pursuitworthiness). Though this advances the field, it produces a different harm: it stagnates our understanding of pursuitworthiness at local minima.

The virtue-economic account acknowledges emerging pursuitworthiness through deliberative judgment. As Duerr and Fischer note, “applying the virtue matrix [...] is a subtle business, not amenable to reduction to rigid rules” (Duerr & Fischer, 2025). In other words, while ‘prefer fruitful experiments’ is established, recognizing fruitfulness in a breakthrough proposal may be just a feeling. This represents progress: unlike the apokritic model, this account recognizes that evaluation involves irreducible emerging forms.

However, this creates differential empowerment when evaluators’ and researchers’ emerging practice diverges. Scientists at the forefront, interacting causally with objects of pursuit, may feel connections between experimental features and fertility that remain opaque. Even when proposals provide all articulable reasons, evaluators using identical virtue matrices may see only incoherence where experimenters feel fruitfulness. Moreover, researchers pursuing breakthroughs often proceed despite, or by suspending established coherence requirements (Sulik et al., 2025). The model handles rational disagreement about established virtue weightings through reticulation (as explained below), but cannot resolve disagreement rooted in emergent pursuitworthiness, where recognizing virtue instantiation depends on unshared feelings. Consequently, acknowledging emergent rationality is insufficient: we still cannot distinguish wrong feelings from feelings-as-not-yet-judgments, and breakthrough experimenters are disadvantaged because their feelings carry less weight than the evaluator’s lack thereof.

Triple reticulation exemplifies the rational explication of emerging intuitions into justified norms. The process seeks reflective equilibrium among three sets: a weighted virtue matrix, a stock of exemplars (paradigmatic scientific achievements), and evaluator’s broader meta-scientific outlook (philosophical and methodological commitments). However, by demanding convergence among *existing* elements, it blocks breakthrough experiments that would introduce *new* ones—new exemplars guided by not-yet-explicated virtues or presupposing novel meta-scientific commitments. By blocking such breakthroughs, it stagnates our understanding of pursuitworthiness at local minima: we cannot learn new standards from experiments we never try. Triple reticulation remains a powerful tool for explication of current emerging pursuitworthiness into established forms, but blocks the development of new emerging elements.

One might object that these accounts were never intended for literal implementation in resource allocation. I agree; their primary aim was understanding pursuitworthiness, not guiding institutional policy. However, the conclusion of this analysis is also general: any account addressing only fragments of the pursuitworthiness landscape will generate dilemma harms through differential empowerment. The next section demonstrates that these harms in scientific practice rest on similar grounds.

4. The Escalating Dilemma Harms

The first horn of the pursuitworthiness dilemma is the conservative bias: by applying established standards consistently, we systematically exclude breakthrough investigations. Critics might argue that the harms of this bias are overstated. They offer two main defenses: first, that we can mitigate them through targeted strategies (4.1); and second, that the gains outweigh the costs, since the alternative—the pain of funding endless failed breakthroughs—is far worse (4.2). I show that both objections underestimate the conservative bias in practice.

4.1. The Illusion of Mitigation

Fischer's (2024) no-lose theorem justifies experiments meeting three conditions: clear realizability, mutually exclusive and exhaustive outcomes, and significant epistemic gain from every possible result. The LHC's Higgs boson search illustrates this: TeV-scale experiments would either observe the particle, confirming the Standard Model, or exclude it, necessitating new physics. Fischer (2024) acknowledges that epistemic gain assessment remains contentious, foreshadowing his virtue-economic account's inclusion of deliberative judgment. For routine zones, no-lose theorems constitute rational pursuitworthiness criteria. However, this not only leaves breakthrough attempts disadvantaged, but incentivizes second-level 'brain drain': researchers abandon emerging insights in chaotic zones for fundable, established directions (Franzoni & Stephan, 2023).

Shaw (2022) proposes mitigating conservative bias via 'luxury science' for competent researchers. Providing evidence that the success of established predictions decays rapidly, he suggests strict norms should apply only to urgent, predictable research. This aligns with our analysis: established pursuitworthiness effectively evaluates only the 'viable here-and-now' routine pursuits. Yet this commits the selection bias: we cannot know if excluded research would yield greater success, even in the short term. For long-term pursuits, Shaw's solution shifts the evaluation from the object of pursuit to the pursuer: by giving luxury funds to "respectable researchers, we can be less worried that resources will systematically go to waste". Shaw acknowledges potential differential empowerment, which I clarify below.

Empowering respectable emergent rationality disadvantages those whose emergent rationality has not yet proven itself. Azoulay et al. (2019) provide evidence that scientific fields experience surges of innovation following dominant researchers' deaths. This 'funeral effect' reveals how senior scientists controlling tenure, funding, and publication don't merely apply explicit norms but enforce decade-old emergent rationality. 'Luxury science' cannot escape this: competent track records reflect success under previous rationality, disadvantaging researchers whose opaque feelings point toward not-yet-conceivable transformations. This systematically excludes contemporary emerging pursuitworthiness—yesterday's breakthroughs become today's barriers.

Lottery-based funding (Avin, 2019) attempts to escape conservative bias through randomization. Yet differential empowerment persists: proposals must pass initial screening, disadvantaging investigations furthest from established pursuitworthiness. On the other hand, reducing entry criteria to include radically novel proposals elevates the permissive bias, increasing waste and embracing nihilism. Therefore, lotteries merely trade the pain of one horn of the dilemma for the other.

4.2 Most breakthrough attempts fail anyway

The strongest defense of conservative bias argues that its cost is negligible compared to the pain of the dilemma's second horn. Researchers prefer routine investigations because 'pursuing innovation is a gamble, without enough payoff, on average, to justify the risk' (Foster et al., 2015). This seemingly justifies advancing incrementally over chasing phantoms. Let us explicate this intuition via a model balancing success chance, epistemic gain, and cost. Assume a simplified specification: established attempts succeed 99% of the time, while breakthroughs succeed 1% but yield 99-fold greater epistemic gain, with equal costs. In such a case, rational allocation of finite resources would require a 50/50 split. The critical questions become: what is the true specification, what do we observe in practice, and how do they relate?

Rzhetsky et al. (2015) provide evidence of a conflict between actual scientific practice and rational standards for experiment selection. Descriptively, modeling 30 years of biomedical literature reveals that researchers follow conservative strategies: focusing on well-studied molecules and immediate neighborhoods. Normatively, however, their simulations reveal that efficient collective discovery requires pursuing distant or disconnected combinations. Though harder to conceive and likely to fail, their success compounds, since bridging disconnected regions facilitates further search. In addition, this misalignment worsens as fields mature because optimal strategies demand risky exploration precisely when practice trends toward safety. Conservative strategies likely also serve individual career needs (steady publication output) over collective epistemic good, a tension we revisit in the final section.

A sophisticated defense might argue that transformative breakthroughs succeed regardless of opposition (Galileo succeeded, after all). Suggesting conservative filters screen the 99% waste while the vital 1% breaks through anyway, is overoptimistic. Barry Marshall's *H. pylori* discovery required self-experimentation to circumvent ethical prohibitions (Thagard, 1998). Stanley Prusiner's 'heretical' prion research survived 'permanent skepticism' only through private funding that bypassed peer review (Aguzzi, 2009; Liberski, 2004). Barbara McClintock worked in isolation for decades because geneticists deemed her 'jumping genes' work 'incomprehensible' (Comfort, 2001). Success required risking death, private wealth, or professional exile—circumstances that bypassed evaluation entirely. We cannot know how many breakthrough attempts failed for lack of such extreme measures.

Moreover, conservative filtering does not randomly exclude radical attempts—it systematically targets the most transformative ones. Siler et al.'s (2015) analysis of 808 manuscripts submitted to elite medical journals revealed that editors rejected the 14 papers that eventually became the most highly cited papers, with 12 of them desk-rejected as 'unworthy of peer review'. Similarly, evaluators systematically score novel proposals lower, even controlling for quality (Boudreau et al., 2016). Veugelers et al. (2022) found that European Research Council (ERC) grants explicitly designed for high-risk/high-gain pioneering research, actually select against applicants with histories of novel work, while Bromham et al. (2016) demonstrated lower funding probabilities for interdisciplinary proposals. In my framework, this is expected: the more a pursuit challenges established norms, the more harshly it will be judged. That's why they need to break through.

Conservative bias intensifies through compounding feedback loops. Park et al. (2023) demonstrate that papers and patents have become less disruptive over time, while Bhattacharya & Packalen (2020) show funding designed for innovation becoming increasingly conservative. Veugelers et al. (2022) found unsuccessful ERC applicants reduce their highly novel research attempts. Scientists report strategic shifts toward mainstream topics to increase funding chances (Franzoni & Stephan, 2023), illustrating how evaluation criteria actively reshape epistemic practices rather than merely measuring them (de Rijcke et al., 2016). Lane et al. (2022) demonstrate how cognitive negativity bias infects peer review: exposing evaluators to low scores triggers disproportionate criticism, dragging down subsequent evaluations far more than high scores raise them. These dynamics create a vicious cycle: conservative evaluation discourages novel research, driving scientists to pursue safer projects, resulting in fewer breakthrough attempts and thereby stagnating our understanding of pursuitworthiness.

The final intensifying mechanism is anti-failure bias: prioritizing success ratios over epistemic gain. Luukkonen (2012) showed that the ERC evaluation process structurally makes feasibility more impactful than groundbreaking potential. The systematic suppression of failure prevents learning from it; publishing failures would accelerate discovery by mapping dead ends (Rzhetsky et al., 2015). If failures became valuable contributions to collective understanding rather than career-ending embarrassments, the entire calculus would shift.

The pursuitworthiness dilemma is not a mere philosophical puzzle but an intensifying scientific crisis. Current practice is not just failing to support breakthroughs; it is evolving to prevent them. This demands a fundamental shift in how we evaluate breakthrough attempts.

5. Foundational transformative norms for funding breakthrough experiments

I argue that we can escape this dilemma and achieve complementary advancement through transformative pursuitworthiness standards. This approach shifts the evaluative focus for breakthrough experiments from *the promise of the idea* to *the promise of the pursuit itself*.

Transformative standards are tailor-made for what breakthroughs do: they assess quality of the process by which a researcher attempts to generate new norms from opaque intuitions. Paul (2014) argues that, while we cannot rationally prefer the outcomes of a transformative experience that changes our preferences, we may choose based on the desirability of the novel preferences themselves. Similarly, we evaluate breakthrough pursuits on their transformative promise to generate new standards. Since the rationality of transformative pursuitworthiness lies in justifying how explicated standards better achieve given ends, its successful application develops established pursuitworthiness itself. Instead of competing with established standards, this approach operates in a complementary space. Evaluating the explication efforts of breakthrough investigations independently avoids the conservative bias while maintaining established standards where applicable. The result is efficiency preserved where appropriate, transformative breakthroughs enabled, and continuous development of our understanding of pursuitworthiness. To realize this, we require transformative standards.

I propose the foundational transformative norm: prefer pursuits that systematically document their emerging practice and their ongoing attempts at its explication and justification. For example, transparently recording metacognitive reflections (e.g., ‘this enzyme feels promising’), tracking pattern recognition (‘successful cases share property X’), and documenting tentative articulations (‘perhaps X indicates success because Y’). Critically, this norm requires capturing not just the intuition or failed justifications, but the transformative practice itself: activities refining those feelings. This transparency transforms private hunches and their development into collective resources, providing the data necessary to evaluate which explication practices work: evidence that empowers the creation of future transformative standards.

Comparison with similar examples clarifies this norm. Kepler disclosed mistaken paths and shifting intuitions in his books, allowing readers to trace evolution of his reasoning. However, today’s publication volume makes such transparency in final papers counterproductive. Contemporary practices like preregistration and registered reports document the state of pursuit before testing, primarily to prevent questionable research practices (Chambers & Tzavella, 2022; Lakens et al., 2024). Proposals exist to extend these to exploratory research—modeled after Victorian explorers who logged every route deviation—making pursuit decisions traceable (Dirnagl, 2020; DeCoster et al., 2015). Similarly, the open lab notebook movement makes incremental results and methodological details public (Schapira et al., 2019; Lang et al., 2011). The shared principle is that whenever new data appear or a decision is made, a record should exist. Yet these approaches rarely document the rationale behind choosing one path over another, nor do they justify the overarching decision-making strategies. This omission likely occurs because the emergent pursuitworthiness driving these decisions is currently deemed unfit for scientific communication.

Unsurprisingly, the practice most closely resembling our proposal emerged in qualitative social sciences, where the researcher’s subjective role is acknowledged as

indispensable. The concept of audit trails (Halpern, 1983) encourages researchers to document the reasons for decisions and the path taken, while maintaining a reflexive journal of evolving thoughts (Rolfe, 2006; Carcary, 2020). Such documentation *facilitates identifying emerging patterns, eliminating non-epistemic influences, and catalyzing transformative explication*. Qualitative researchers deem this documentation necessary to evaluate their methodology where the researcher's subjective, transformative role is inseparable from the inquiry. Quantitative and experimental researchers argue adopting these reflexive practices to counteract structural biases (Acem et al., 2025; Jamieson et al., 2023). My aim is more general: I propose using this documentation universally to create a collective space for developing new pursuitworthiness standards.

This norm's novelty lies in explicitly including scientists' intuitive judgments, affective responses, and evolving sense of direction—those 'messy' yet indispensable elements typically excluded from 'objective' accounts. Whereas other practices document articulated reasons, this norm calls for recording feelings even before reasons can be explicated. Furthermore, its scope is maximal, applying across formal, natural, social, and humanistic disciplines. In essence, it demands *documenting the drivers behind deciding on one path rather than another*, even if that is not yet a reason.

The documentation norm could ground an independent funding mechanism for breakthrough research. Instead of selecting for idea promise (filtering out radical breakthroughs), this stream would evaluate solely on explication rigor. Supported researchers would maintain detailed documentation: recording discussions, decisions, and iterative reasoning to make their ideas' evolution empirically traceable. Competition would thus shift to epistemic accountability: *who can make the process of pursuit most intelligible*. On the other hand, low-effort documentation could become an additional evaluative dimension across existing pursuitworthiness decisions. Beyond digital repositories already mentioned, journals could establish 'explication narrative' formats, valorizing well-documented pursuits regardless of success, or encouraging 'pursuit histories' as supplementary material. Similarly, hiring standards could include commitments to this norm (with descriptions of such practice made public). Collectively, these would generate vast new data on what makes pursuits worthy.

Direct engagement with such evidence offers immediate benefits to researchers working on related problems. Documenting small successes and failures in real time, such as effective temperatures or unreliable reagents, allows others to avoid unproductive paths and build on partial progress. Moreover, when documentation also records the underlying intuitions that motivated particular choices, the evolution of feelings, and nascent patterns, integration occurs at the level of process, not just results. Researchers gain access to the unfolding rationality that guided earlier attempts, inheriting conceptual and affective scaffolding rather than isolated checkpoints. This enables the community to recognize misleading intuitions earlier and refine promising ones more effectively.

Moreover, this evidence enables the further development of pursuitworthiness accounts. While philosophers have long drawn insights from historical materials like

Darwin's or Maxwell's notebooks (Gruber & Wallace, 2001; Nersessian, 2008), such post-hoc reconstructions suffer from limited generalizability based on few, possibly cherry-picked cases (Pitt, 2001). Systematic documentation of pursuitworthiness practice across disciplines would overcome these limitations. It would provide meta-scientists—philosophers (Leonelli, 2019, 2023; Miłkowski & Nowakowski, 2025), methodologists (Vazire, 2018; Mayo, 2018), psychologists (Aczel, 2024), and sociologists (Evans & Foster, 2011), among others—with the data necessary to construct empirically grounded, context-sensitive standards based on actual practice rather than idealized reconstructions.

This evidence would serve as both inspiration and evaluative benchmark. Analyzing documentation of successful and failed pursuits would allow us to discover standards for recognizing misleading or promising feelings earlier, and for identifying which transformative practices best evolve them into justified forms. Ultimately, advancing our understanding of pursuitworthiness requires explicit justifications of these standards as suitable means for given ends, uniting the distinct strengths of meta-scientific fields: philosophers contributing conceptual analysis, cognitive scientists modeling systems, and sociologists analyzing institutional structures. In this way, the documented practice of pursuit becomes the foundation for a robust, complete account of pursuitworthiness. The following examples illustrate the kinds of transformative norms such a process could uncover, to spark the imagination.

Consider a first potential norm: *prefer pursuits that, in the face of persistent failure, fundamentally alter core assumptions rather than iteratively refine existing frameworks*. Kepler's shift from perfect Platonic solids to 'imperfect' ellipses exemplifies this. This very paper embodies this norm: in response to the persistent dilemma, I shift the evaluative focus from the object to the process of pursuit. Practically, funding agencies should prioritize proposals altering fundamental assumptions in research areas of diminishing returns.

Consider a second potential norm: *prefer pursuits that develop intuitions regarding new evidence types or qualities, especially amid persistent failure*. Kepler's breakthrough depended on Tycho Brahe's unprecedented data precision. This aligns with our earlier analysis of causal interaction as epistemically generative. Pursuits that invest in new measurement techniques, access unobservable phenomena, or generate novel data types demonstrate transformative potential by expanding the evidential basis itself. Notably, the foundational documentation norm embodies this principle, generating novel evidence: the description of practice that makes research pursuitworthy itself.

These norms illuminate our departure from Lakatos's treatment of research programs. Where Lakatos deemed 'degenerating' programs (those solving fewer problems) unworthy of pursuit, transformative pursuitworthiness recognizes apparent *degeneration as a signal to seek different standards*. This illustrates the contextual and dynamic nature of choosing pursuitworthiness standards. While previous examples focus on recognizing

emergent pursuitworthiness worth explicating, the following norm illustrates distinguishing better transformative pursuitworthiness practice from worse.

Consider an explication norm: *prefer intertwined pursuits over isolated ones*. Darwin's development of evolutionary theory exemplifies what Gruber termed a 'network of enterprise': multiple, simultaneous lines of inquiry where progress in one domain catalyzes breakthroughs in others. Barnacle classification informed his view of variation, geological studies shaped his temporal intuitions, and plant breeding clarified selection. Institutionally, this suggests prioritizing networked pursuits; for example, funding laboratory networks where distinct but related investigations regularly exchange insights. Practices such as cross-laboratory meetings or rotation programs can facilitate this fertilization. The foundational documentation norm makes this possible, as transparently recording the process allows researchers to effectively intertwine their pursuits.

Recall that pursuitworthiness operates across diverse scales, mediums, and roles: from a reviewer suggesting a direction, through oral conference discussions, to intergovernmental decisions on building a larger hadron collider. As a result, documentation forms and standards must vary. At a conference, for instance, an audience might use 'confusion', 'worry', or 'promise' buttons to provide real-time affective feedback. This allows the presenter to gauge the room's collective feeling—what to clarify, defend, or develop—beyond the few able to ask explicit questions. Moreover, because initial emergent pursuitworthiness is highly embodied, embedded, and dynamic, reflexive journals or recordings of informal discussions may constitute the only possible evidence of these early stages. Accordingly, standards here must be less strict, taking the form of flexible guidelines rather than rigid rules. Finally, for external evaluators like mentors or reviewers, we require standards for assessing and nurturing transformative potential of others' hunches.

Transformative pursuitworthiness aligns naturally with accounts locating standards in the pursuing subject, whether individual or collective. This directly connects to Shaw's (2022) insight: if long-term gain depends on competent pursuers, then developing scientists' explication competencies is crucial. Training programs could foster these skills through metacognitive awareness exercises and collaborative sense-making workshops (Gardiner, 2020). At the collective level, shaping research culture supports these competencies. Candiotto (2022) argues that emotions become truth-motivating only in epistemic cultures that value and embed habits of participatory inquiry. Her point is that specific environments—comprising material arrangements (e.g., circular seating) and practices (e.g., question-oriented discussion)—scaffold agents' efforts. By cultivating both these explication competencies and the habits within the research environment, we create a system more adept at transforming the emerging into the established.

I now address three kinds of pressing objections regarding the foundational norm: its effectiveness, its absence in current practice, and its systemic consequences.

First: Does documenting feelings waste resources or stifle emergence? Critics might argue that recording intuitions diverts effort from actual research. However, the minimal version of this norm simply requires making existing private notes public; researchers

already track these feelings, and I merely advocate sharing them. Extended documentation applies only to research funded specifically on this criterion. One might worry that heavy documentation burdens could stifle the possibly crucial spontaneity of emergent practice. Here, digitalization and AI integration offer hope for making documentation increasingly frictionless, ensuring that the transparency mechanism does not choke the process it aims to reveal. Nevertheless, trading short-term, task-level efficiency (time spent pipetting) for long-term, systemic efficiency (generating the data needed to escape the conservative bias loop) is rational.

Second: If documenting emerging rationality is so valuable, why isn't it already a norm? Three intertwined factors explain this. First, the regulative ideal of objectivity systematically eliminates subjective elements from public communication, viewing them as threats to science's rational image. Second, we lack a fundamental understanding of everyday scientific practice; scientists perform competently without articulating how, learning first via imitation-based apprenticeship rather than explicit instruction. Third, the early twentieth-century philosophy of science, *nota bene* responding to revolutionary breakthroughs, focused exclusively on knowledge products rather than the processes that create them. Because we rightly aim at objective products, we have historically marginalized the subjective processes necessary to produce them.

Third: Won't researchers fear being scooped? Indeed, frontier researchers may be reluctant to publicize intuitions, fearing that competitors will leverage them without comparable investment; a newcomer might surpass the originator by building on documented intuitions with a fresh perspective. This reluctance stems from Merton's 'priority rule': only the first discoverer receives recognition. Originally, this norm served efficiency by discouraging redundancy of latecomers and encouraging urgency at the frontiers. However, in today's hypercompetitive system, it operates destructively, discouraging process-sharing and open collaboration. While shifting rewards from individuals to communities (Tiokhin et al., 2024) offers one transformative solution, this highlights a fundamental conflict between collective and individual given ends. Until this divergence is resolved—until what benefits truth also benefits the researcher—implementation of any epistemic norm risks producing unintended consequences.

6. Conclusions

I hope to have convinced the reader that the development of a complete transformative pursuitworthiness account is a philosophical project worth pursuing.

To describe pursuitworthiness at the forefront of science I introduced the triadic framework (Table 1) and the practice landscape (Figure 1). The incompleteness of this map charts the philosophical work ahead.

To ground this framework, I advanced the plausibility of emergent pursuitworthiness rationality through the cognitive science of metacognitive feelings. This causal account suggests that such feelings possess nascent rationality because they encode

object-derived information unavailable to established reasoning alone. Most breakthroughs are not irrational leaps in the dark; they are guided by metacognitive feelings carrying objective feedback from causal interactions at the scientific frontier. This internal rationality implies that scientists should follow their emergent intuitions, but only if they treat them as half-baked hypotheses awaiting falsification and refinement.

I defended the rationality of transformative pursuitworthiness by demonstrating that norms like ‘prefer simplicity’ gain force over ‘prefer what feels easy’ because they better achieve the given end of efficiency. This second-order instrumental rationality aligns with Fleisher’s (2022) proposition that inquiry norms must promote successful inquiry. Rather than in conflict with traditional epistemic norms (Friedman, 2020), transformative standards function as meta-norms, guiding epistemic complexities of inquiry. And while Thorstad (2022) identifies the inevitable non-epistemic factors shaping inquiry, transformative practice aims precisely to isolate the epistemic from the non-epistemic, crystallizing the former into justified norms.

Partial pursuitworthiness accounts and current funding structures do not merely limit breakthrough research; they systematically penalize its foundational mechanisms. Demanding explicit justification for opaque feelings condemns emergent ideas for lacking what cannot yet exist. Evaluators lacking causal interaction with the object of inquiry cannot grasp the intuitions of breakthrough proposals. Furthermore, explication through coherence within pursuitworthiness of today, or empowering historically successful researchers, disfavors the rationality of tomorrow. Even funding lotteries merely trade conservative bias for permissive scientific waste. Worse, this conservative filtering systematically targets the most transformative attempts. The conflict between rational exploratory strategies and observed conservative practice intensifies through compounding feedback loops—such as brain drain and anti-failure bias—trap science into a self-reinforcing cycle against the breakthroughs.

To escape these harms, I propose a foundational transformative norm: prefer pursuits that systematically document their emerging pursuitworthiness and explication efforts. Admittedly, fully operationalizing this norm remains a task for future research, aligning with growing calls from meta-scientists for experimental funding designs (Wilsdon et al., 2015; Bendiscioli et al., 2022). Yet, institutionalizing this norm would provide an independent funding mechanism for breakthrough experiments—one that demands rigorous epistemic accountability without requiring impossible foresight, protecting early-stage breakthroughs from their lethal evaluation through already established standards. For scientists, it provides a rationally justified sanctuary to pursue breakthroughs. For philosophers and meta-scientists, it transforms the ‘chaos of feeling’ into a tangible, collective resource, generating the vast, previously inaccessible empirical data required to understand how pursuits actually work. While I do not propose standards to predict successful explications, demanding the documentation of the transformative journey provides the empirical foundation to develop them. This paper does not solve the

recognition problem; it merely shifts its dimension from evaluating the object of pursuit to evaluating the process.

Likewise, this framework does not eradicate differential empowerment. The documentation norm risks privileging researchers with superior metacognitive awareness, pursuits that are inherently easier to document, or institutions with the resources to support such reflexivity. Yet, this recursive challenge reveals a deeper insight that addresses Feyerabend's critique: any partial pursuitworthiness account generates some form of differential empowerment. Therefore, analyzing scientific gain versus waste, alongside the advancement versus stagnation of pursuitworthiness, must become an integral part of evaluating pursuitworthiness standards themselves. This framework does not ultimately resolve the dilemma, but it makes its dynamics visible and addressable, transforming a paralyzing paradox into a productive tension that continuously drives complementary advancement of scientific practice and our understanding of pursuitworthiness itself.

References

- Acem, E., Aczel, B., Albayrak, N., Brown, N. J., Dudda, L. A., Elsherif, M. M., ... & Zubaly, B. (2025). Why I declare a conflict of interest and you should not. *Theory and Society*, 54(6), 1137-1172.
- Ackerman, R., & Thompson, V. A. (2017). Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617.
<https://doi.org/10.1016/j.tics.2017.05.004>
- Aczel, B. (2024). Why we need a 'psychology of science'. *Nature Human Behaviour*, 8(1), 4-5.
- Amrhein, V., & Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, 2, 4–4.
- Arango-Muñoz, S. (2014). The nature of epistemic feelings. *Philosophical Psychology*, 27(2), 193–211. <https://doi.org/10.1080/09515089.2012.732002>
- Avin, S. (2019). Mavericks and lotteries. *Studies in History and Philosophy of Science Part A*, 76, 13–23. <https://doi.org/10.1016/j.shpsa.2018.11.006>
- Azoulay, P., Fons-Rosen, C., & Zivin, J. S. G. (2019). Does science advance one funeral at a time?. *American Economic Review*, 109(8), 2889-2920.
- Aguzzi, A. (2009). Beyond the prion principle. *Nature*, 459(7249), 924-925.
- Bendiscioli, S., Firpo, T., Biosca, A. B., Czibor, E., Garfinkel, M. S., Stafford, T., Wilsdon, J & Woods, H. B. (2022). *The Experimental Research Funder's Handbook*. Research on Research Institute.
- Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Management Science*, 62(10), 2765–2783.
<https://doi.org/10.1287/mnsc.2015.2285>

- Brading, K., & Castellani, E. (2003). *Symmetries in physics: Philosophical reflections*. Cambridge University Press.
- Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609), 684-687.
- Bhattacharya, J., & Packalen, M. (2020). *Stagnation and scientific incentives* (No. w26752). National Bureau of Economic Research.
- Candiotto, L. (2022). Epistemic emotions and co-inquiry: A situated approach. *Topoi*, 41(5), 839-848.
- Carcary, M. (2020). The research audit trail: Methodological guidance for application in practice. *Electronic Journal of Business Research Methods*, 18(2), pp166-177.
- Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature human behaviour*, 6(1), 29-42.
- Chang, H. (2012). *Is water H₂O?: Evidence, realism and pluralism* (Vol. 293). Springer Science & Business Media.
- Collins, H. (2010). Tacit and explicit knowledge. In *Tacit and explicit knowledge*. University of Chicago press.
- Comfort, N. C. (2001). *The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control*. Harvard University Press. <https://doi.org/10.4159/9780674029828>
- Currie, A. (2023). Epistemic engagement, aesthetic value, and scientific practice. *The British Journal for the Philosophy of Science*, 74(2), 313-334.
- Daston, L., & Galison, P. L. (2021). *Objectivity*. Princeton University Press.
- DeCoster, J., Sparks, E. A., Sparks, J. C., Sparks, G. G., & Sparks, C. W. (2015). Opportunistic biases: Their origins, effects, and an integrated solution. *American Psychologist*, 70(6), 499.
- de Rijcke, S., Wouters, P. F., Rushforth, A. D., & Franssen, T. P. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>
- Doudna, J., & Sternberg, S. (2017). *A crack in creation: The new power to control evolution*. Random House.
- Duerr, P. M. & Fischer, E. (2025): Rationally warranted promise: the virtue-economic account of pursuit-worthiness. *Synthese*, 206(2), 68.
- Dirnagl, U. (2020). Preregistration of exploratory research: Learning from the golden age of discovery. *PLoS biology*, 18(3), e3000690.
- DiMarco, M., & Khalifa, K. (2022). Sins of inquiry: How to criticize scientific pursuits. *Studies in History and Philosophy of Science*, 92, 86-96.
- Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. Verso Books.
- Firestein, S. (2012). *Ignorance: How it drives science*. Oxford University Press.
- Fischer, E. (2024). No-lose theorems and the pursuitworthiness of experiments. Preprint. <https://philsci-archive.pitt.edu/23856/>
- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical*

- Society*, 98(1), 39–82.
- Fleck, L. (1981). *Genesis and development of a scientific fact*. University of Chicago Press.
- Fleisher, W. (2022). Pursuit and inquisitive reasons. *Studies in History and Philosophy of Science*, 94, 17–30.
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, 80(5), 875–908.
<https://doi.org/10.1177/0003122415601618>
- Franklin, A. (1989). *The neglect of experiment*. Cambridge University Press.
- Franzoni, C., & Stephan, P. (2023). Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals. *Research Policy*, 52(3), 104706.
- Friedman, J. (2020). The epistemic and the zetetic. *Philosophical review*, 129(4), 501-536.
- Evans, J. A., & Foster, J. G. (2011). Metaknowledge. *Science*, 331(6018), 721-725.
- Galison, P. (1987). *How Experiments End*. University of Chicago Press.
- Gardiner, P. (2020). Learning to think together: Creativity, interdisciplinary collaboration and epistemic control. *Thinking Skills and Creativity*, 38, 100749.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587–606.
- Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Penguin.
- Gottlieb, S., Keltner, D., & Lombrozo, T. (2018). Awe as a Scientific Emotion. *Cognitive Science*, 42(6), 2081–2094.
- Gruber, H. E., & Wallace, D. B. (2001). Creative work: The case of Charles Darwin. *American Psychologist*, 56(4), 346.
- Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.
- Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge university press.
- Halpern, E. S. (1983). *Auditing naturalistic inquiries: The development and application of a model*. Indiana University.
- Jamieson, M. K., Govaart, G. H., & Pownall, M. (2023). Reflexivity in quantitative research: A rationale and beginner's guide. *Social and personality psychology compass*, 17(4), e12735.
- Kaptchuk, T. J. (1998). Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine*, 72(3), 389–433.
- Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford University Press.
- Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. *Handbook of Metamemory and Memory*, 117–135.
- Kozlov, A. (2023). Scientific experiments beyond surprise and beauty. *European Journal for Philosophy of Science*, 13(3). <https://doi.org/10.1007/s13194-023-00536-7>
- Leonelli, S. (2019). *Data-centric biology: A philosophical study*. University of Chicago Press.

- Leonelli, S. (2023). *Philosophy of open science*. Cambridge University Press.
- Lakens, D., Mesquida, C., Rasti, S., & Ditroilo, M. (2024). The benefits of preregistration and Registered Reports. *Evidence-Based Toxicology*, 2(1), 2376046.
- Lang, A., Bradley, J. C., Koch, S., & Neylon, C. (2011). Collaboration Using Open Notebook Science in Academia. In S. Ekins, M. A. Hupcey, & A. J. Williams (Eds.), *Collaborative Computational Technologies for Biomedical Research* (pp. 425–452). John Wiley & Sons.
- Lander, E. S. (2016). The heroes of CRISPR. *Cell*, 164(1), 18–28.
- Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth* (Vol. 282). Univ of California Press.
- Lane, J. N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E. C., & Lakhani, K. R. (2022). Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management science*, 68(6), 4478–4495.
- Liberski, P. P., & Ironside, J. W. (2004). An outline of the neuropathology of transmissible spongiform encephalopathies (prion diseases). *Folia neuropathologica*, 42, 39–58.
- Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research evaluation*, 21(1), 48–60.
- Li, M., & Vitányi, P. (2019). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer International Publishing.
<https://doi.org/10.1007/978-3-030-11298-1>
- Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press.
- McAllister, J. W. (1999). *Beauty and revolution in science*. Cornell University Press.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
- Milkowski, M., & Nowakowski, P. (2025). Tracking norms: a plea for a normative digital philosophy of science. *Synthese*, 206(3), 120.
- Miyata-Sturm, M. E. (2024). Aesthetic Considerations in the Development of Plate Tectonics. *Studies in History and Philosophy of Science*, 108, 1–9.
- Mooij, J. M., Magliacane, S., & Claassen, T. (2020). Joint causal inference from multiple contexts. *Journal of machine learning research*, 21(99), 1–108.
- Morton, A. (2009). Epistemic Emotions. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion* (pp. 385–400). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199235018.003.0018>
- Nerantzaki, K., Efklides, A., & Metallidou, P. (2021). Epistemic emotions: Cognitive underpinnings and relations with metacognitive feelings. *New Ideas in Psychology*, 63, 100904. <https://doi.org/10.1016/j.newideapsych.2021.100904>
- Nersessian, N. J. (2008). *Creating scientific concepts*. MIT Press.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
<https://doi.org/10.1073/pnas.1708274114>
- Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive

- over time. *Nature*, 613(7942), 138-144.
- Paul, L. A. (2014). *Transformative experience*. OUP Oxford.
- Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.
- Pitt, J. C. (2001). The dilemma of case studies: Toward a Heraclitian philosophy of science. *Perspectives on Science*, 9(4), 373-382.
- Polanyi, M. (1962). Tacit Knowing: Its Bearing on Some Problems of Philosophy. *Reviews of Modern Physics*, 34(4), 601-616. <https://doi.org/10.1103/revmodphys.34.601>
- Shan, Y. (2020). Promisingness in Theory Choice, *Boston Studies in the Philosophy and History of Science* (pp. 177-192). Springer International Publishing. https://doi.org/10.1007/978-3-030-50617-9_10
- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford University Press.
- Radder, H. (2003). *The philosophy of scientific experimentation*. University of Pittsburgh Press.
- Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of advanced nursing*, 53(3), 304-310.
- Rzhetsky, A., Foster, J. G., Foster, I. T., & Evans, J. A. (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences*, 112(47), 14569-14574. <https://doi.org/10.1073/pnas.1509757112>
- Schapira, M., Open Lab Notebook Consortium, & Harding, R. J. (2019). Open laboratory notebooks: Good for science, good for society, good for scientists. *F1000Research*, 8, 87. <https://doi.org/10.12688/f1000research.17710.2>
- Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, 31(1), e2295. <https://doi.org/10.1002/icd.2295>
- Schwartz, M. A. (2008). The importance of stupidity in scientific research. *Journal of Cell Science*, 121(11), 1771-1771. <https://doi.org/10.1242/jcs.033340>
- Seselja, D., Kosolosky, L., & Straßer, C. (2012). The rationality of scientific reasoning in the context of pursuit: Drawing appropriate distinctions. *Philosophica*, 86, 51-82.
- Shaw, J. (2022). On the very idea of pursuitworthiness. *Studies in History and Philosophy of Science*, 91, 103-112.
- Shea, N. (2023). Metacognition of Inferential Transitions. *Journal of Philosophy*. 121(11), 597-627.
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4), 186-193. <https://doi.org/10.1016/j.tics.2014.01.006>
- Schurz, G., & Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Topics in Cognitive Science*, 11(1), 7-36.
- Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, 112(2), 360-365. <https://doi.org/10.1073/pnas.1418218112>
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of*

- Economics*, 99–118.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. <https://doi.org/10.1098/rsos.160384>
- Spirtes, P., Glymour, C., & Scheines, R. (2001). Causation, Prediction, and Search, 2nd Edition. In *MIT Press Books* (Vol. 1). The MIT Press.
<https://ideas.repec.org/b/mtp/titles/0262194406.html>
- Stacey, R. D. (1996). *Complexity and creativity in organizations*. Berrett-Koehler Publishers.
- Stacey, R. D. (1992). *Managing the unknowable: Strategic boundaries between order and chaos in organizations*. John Wiley & Sons.
- Stacey, R. D. (2007). *Strategic management and organisational dynamics: The challenge of complexity to ways of thinking about organisations*. Pearson education.
- Steinle, F. (2002). Experiments in history and philosophy of science. *Perspectives on Science*, 10(4), 408–432.
- Sulik, J., Rim, N., Pontikes, E., Evans, J., & Lupyan, G. (2025). Differences in psychologists' cognitive traits are associated with scientific divides. *Nature Human Behaviour*, 9(6), 1147–1161. <https://doi.org/10.1038/s41562-025-02153-1>
- Tiokhin, L., Panchanathan, K., Smaldino, P. E., & Lakens, D. (2024). Shifting the level of selection in science. *Perspectives on Psychological Science*, 19(6), 908–920.
- Thagard, P. (1998). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 29(1), 107–136.
- Thorstad, D. (2022). There are no epistemic norms of inquiry. *Synthese*, 200(5), 410.
- Veugelers, R., Wang, J., & Stephan, P. (2022). *Do funding agencies select and enable risky research: Evidence from ERC using novelty as a proxy of risk taking* (No. w30320). National Bureau of Economic Research.
- Valdesolo, P., & Graham, J. (2014). Awe, Uncertainty, and Agency Detection. *Psychological Science*, 25(1), 170–178. <https://doi.org/10.1177/0956797613501884>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13(4), 411–417.
- Vogl, E., Pekrun, R., & Loderer, K. (2021). Epistemic Emotions and Metacognitive Feelings. In D. Moraitou & P. Metallidou (Eds.), *Trends and Prospects in Metacognition Research across the Life Span: A Tribute to Anastasia Efklides* (pp. 41–58). Springer International Publishing. https://doi.org/10.1007/978-3-030-51673-4_3
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., & Kerridge, S. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Higher Education Funding Council for England (HEFCE). <https://doi.org/10.13140/RG.2.1.4929.1363>
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Harvard University Press.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Zimmerman, B. (2001). *Ralph Stacey's agreement & certainty matrix*. Schulich School of Business, York University, Toronto, Canada.