Working Paper
please do not circulate without permission

The Art of the Short Story:

Second-Order Inference and Rhetorical Fit in Strategic Communication

Georg Rilinger & Brad Turner[1]

ABSTRACT

Why do activist short sellers publish sprawling, theatrical reports, even though their profits depend on rapid revaluation and existing theory suggests that focused, conventional signals coordinate investors best? To answer this question, our abductive study combines computational analysis of 830 reports targeting U.S.-listed firms (2008–2024) with 21 practitioner interviews, market data, and audience reception evidence. The corpus reveals that report architecture tracks the evaluative environment: tone reflects house style, but rhetorical breadth rises with evaluative uncertainty and, among broad reports, narrative integration rises with pre-campaign volatility, our indicator of active contestation. Interviews reveal the logic behind these patterns. Sellers anticipate not how "the market" will respond (third-order inference) but how specific audiences will react in sequence (second-order inference). When audiences share an evaluative standard, the two coincide and focused reports suffice. When standards diverge, modular reports make competing standards and their audiences visible; when uncertainty coincides with active contestation, integrated reports bind these lines into an overarching narrative that is easier to grasp and transmit. Both departures sacrifice genre conformity, so sellers reserve them for audiences that focus cannot coordinate. Comparing our explanation—rhetorical fit—with accepted alternatives, we develop a more general account of coordination in valuation processes: the focused signal of focal-point theory is the special case of aligned standards, and rhetorical complexity enables second-order inference where shared standards are absent.

[1] MIT Sloan School of Management

INTRODUCTION

Valuation is the social process through which actors arrive at judgments of worth about goods, services, or producers (Karpik 2010; Lamont 2012; Zuckerman 2012a). Organizations that want to influence valuation processes must contend with at least two problems. First, actors rarely share a single evaluative standard (Gouvard and Durand 2023; Hsu and Hannan 2005). Depending on their position, expertise, and interests, they attend to different facts and weigh them by different criteria (Bourdieu 1993; Cattani, Ferriani, and Allison 2014; Zuckerman 2017). Second, actors rarely judge independently of one another (Botelho 2024; Reilly 2026).

In many settings, acting on an evaluative judgment pays only if enough others do so as well. Joining a revolutionary movement pays only if others join, adopting a network technology only if others adopt, and a short position becomes profitable only if enough investors sell (Chwe 2003). Actors therefore cannot simply follow their own judgment. They must base their decisions on inferences about the judgments of others, who are doing the same (Zuckerman 2004). Organizations that seek to influence such a process face more than a problem of persuasion: they must shape what heterogeneous audiences expect each other to believe and do.

Existing research suggests a baseline expectation for how organizations can meet this challenge: they can anchor mutual expectations with a focal point—a widely perceived signal whose prominence is itself common knowledge (Chwe 2003; Schelling 1960). Presidential announcements, scandals, or collective rituals can synchronize action because they facilitate 'third-order' inference: they help actors anticipate what most others will believe or do (Correll et al. 2017). Strategic communication that serves this function should be focused and conventional. Focus makes a text easier to understand and to pass on, and harder to contest (Bruner 1986; Charles and Kendall 2024; Heath, Bell, and Sternberg 2001). Conforming to a recognized genre

signals credibility because the text can be judged by a standard audiences already share (Cutolo and Ferriani 2024).

Activist short selling is a setting in which this expectation should hold particularly strongly. Short sellers establish positions that become profitable when a company's share price falls and then publish reports intended to persuade other market participants that the company is overvalued or fraudulent (Botelho 2018; Paugam, Stolowy, and Gendron 2021; Zuckerman 2012b). Because a short seller's position becomes profitable only when other investors sell, the report must do more than establish the correctness of an investment thesis. It must help generate a sufficiently broad and coordinated market response – and quickly. We would therefore expect short sellers to organize their evidence around a focused investment thesis and to present it in the voice of professional financial analysis or an established brand.

Yet short sellers frequently depart from this expectation. Fuzzy Panda's 2022 campaign against Fisker, for example, combined more than 16,000 words of financial analysis, anonymous testimony, customer complaints, mockery, moral accusation, and visual spectacle, including a narrative that the founder had engineers "SMUGGLE parts into Finland." Such reports are not outliers. Across our corpus, short sellers frequently publish sprawling and irreverent accounts that combine multiple lines of attack and depart from the conventions of professional investment research. Nor is this simply a matter of house style: the same short seller may publish a restrained, narrowly argued report against one company and an expansive, theatrical report against another. If focused and conventional accounts secure attention and credibility, why would experienced communicators relax precisely these features? Mapping and explaining this variation is the goal of this paper.

We draw on a mixed-methods, abductive study of 830 activist short-selling reports targeting U.S.-listed companies between 2008 and 2024. We combine close reading and computational analysis of the reports with market data, analyst reports, earnings-call transcripts, journalistic coverage, retail-investor forums, and 21 interviews with short sellers and financial journalists. A first step establishes the puzzle by mapping the variation in the rhetorical architecture of short sellers' reports. We identify three distinct dimensions: their rhetorical breadth, the presence of an overarching narrative, and tone. While tone, the intensity of stylistic devices, largely reflects an effort to cultivate a house style, this is not true of the breadth and integration of reports.

To develop an explanation, we turn to interviews that reveal the coordination problem short sellers see themselves as solving. They try to trigger a cascade of trades that moves from algorithms and speculative traders to retail and then institutional investors. Actors early in this sequence do not simply anticipate how "the market" will respond. They anticipate the reactions of specific audiences that follow them. The relevant coordination problem therefore involves second-order inference, not only third-order inference about the market as a whole. The interviews further distinguish two difficulties: uncertainty about which evaluative standards later audiences will apply and active contestation from rival valuations competing for attention.

Returning to the corpus, we find that greater evaluative uncertainty is associated with broader reports. Pre-campaign volatility does not distinguish broad from focused reports, but it is associated with greater narrative integration among broad reports. We then analyze typical cases to discover the strategic logic behind these associations. When evaluative standards are settled, *focused* reports organize a small number of claims around a recognizable investment thesis. In these cases, audiences all judge by the same standard and the problem of second-order inference reduces to the classic problem of third-order inference. A focused and genre-conforming report

enables each reader to infer that most others will find it credible. In contrast, when standards are unsettled, activists use *modular* reports that develop separable lines of attack. These not only address heterogenous standards, but also make these standards — and the audiences attached to them — visible to everyone, thereby facilitating second-order inference. Because these reports are unwieldy, they are harder to understand and pass on, which becomes a problem when the target company is contested and attention is scarce. When evaluative uncertainty coincides with active contestation, *integrated* reports therefore add a narrative throughline that makes this broader case easier to grasp and transmit amid competing accounts. Each departure from the focused style carries risks that literature has identified: breadth exposes weak claims, and integration can be read as manipulative or unprofessional. This suggests a theory of rhetorical fit: short sellers tend to incur those risks when the difficulty of second-order inference demands it.

We make three contributions. First, we show that the rhetorical structure of strategic communication is not just a tool of persuasion, but of coordination. It can facilitate inferences about the beliefs of others. Second, the structure of strategic communication may vary not just in relation to audience preferences, but relative to the coordination problems it is meant to address. While audiences might prefer a focused and conventional text, this text may not be as effective for coordinating interdependent audiences with heterogeneous evaluative standards. Third, we suggest that coordination in financial markets does not depend only on third-order inference about the broader market, but can also include second-order inference about specific other audiences. This can make the coordination problems of arbitrage even harder than commonly assumed.

THEORY

Valuation and Interdependent Audiences

Many valuation processes are characterized by the presence of heterogeneous audiences (Isenring, Durand, and Laamanen 2025; Reilly 2026): different actors evaluate the same object differently. For instance, novice investors buy the stocks that catch their attention while institutions weigh fundamentals (Barber and Odean 2007); venture capitalists welcome an ambiguous category label as a signal of growth potential while consumers penalize it as an incoherent identity (Pontikes 2012). Even evaluators with the same evaluative standards may differ in how they apply them. Some may compare an offering to a category prototype, while others relate it to their own goals, leading to opposite judgments about the same object (Boulongne and Durand 2021; Gouvard and Durand 2023). How to deal with the presence of multiple evaluative standards is therefore a crucial question for many organizations that want to shape valuation processes.

When audiences judge independently of one another, the answer is simple. An organization can mold itself to fit the evaluative standards of one particular audience (Zuckerman 1999), or it can address multiple audiences separately by tailoring a presentation to each audience's evaluative standard: publishers release a literary imprint for critics and prize juries and a mass-market imprint for casual readers, and neither audience needs to know what the other is shown (Hsu and Hannan 2005). But such tailoring can become difficult when audiences do not just apply different evaluative standards but make their own judgments dependent on each other's.

The literature has identified at least three forms of audience interdependence. First, audiences' interests can be opposed such that some might see the very attempt to serve other parts of the market as betrayal: corporate clients tolerate a law firm that also practices family law but defect from one that represents plaintiffs against corporations (Phillips, Turco, and Zuckerman 2013). Second, audiences' tastes can be defined against each other, as when beer enthusiasts devalue craft brands owned by industrial brewers (Carroll and Swaminathan 2000). Third, their

judgments can sequentially influence each other. When a platform makes prior judgments visible to new users, they often adjust their own (Botelho 2024). In all three cases, organizations can specialize on a subset of audiences to respect these interdependencies. A high-status law firm can turn away plaintiffs' work, and a platform can focus on marquee users.

Strategies of tailoring or specialization work only when an organization can profit from the evaluative judgments of individual audiences. Every beer the brewery sells to the mainstream adds to its profits even when the craft beer enthusiasts turn up their noses. But this is not true for all valuation processes. In many cases, an organization's ability to profit requires that the evaluative judgments of different audiences align with one another. A critical mass of heterogeneous audiences must act in concert for the organization to profit.

Speculative bubbles in asset markets make this logic most transparent. If an investor notices that an asset is overpriced relative to its fundamentals, they can only profit from this insight if enough others come to the same conclusion around the same time. Even when everyone applies the same evaluative standard and privately agrees that an asset is overpriced, the inflated price may persist if decisions to sell this asset are not synchronized (Abreu and Brunnermeier 2003). The judgments of others thus matter not because they affect an actor's private evaluative judgment - as they do for the beer enthusiast - but because these other judgments determine whether it pays to act on one's own judgment.

This kind of threshold problem characterizes many valuation processes (Chwe 2003): actors choosing whether to join a revolutionary movement, adopt a network technology, or invest in a startup only stand to benefit when enough others join. Indeed, in many cases it may be beneficial *not* to act if the threshold has not been passed yet - to denounce the movement one secretly

admires, or to ride the financial bubble and "dance" with a dangerous position in the hope of getting out before the music stops (Turco and Zuckerman 2014). Because of this, coordinated action may not occur even when everyone privately agrees with each other. This logic explains why an open secret can persist for years and then suddenly trigger a scandal: publicity exposes a transgression to everyone simultaneously, and everyone now expects everyone else to react (Adut 2005). Yet, publicity alone does not typically resolve the problem. During the private equity bubble of the mid-2000s, a prominent insider publicly warned that prices had lost touch with fundamentals, but the bubble did not burst: the warning told investors what its author believed but not what their peers would do (Turco and Zuckerman 2014). A public signal coordinates only when it resolves each actor's uncertainty about the other audiences.

Correll et al. (2017) distinguish three orders of inference problems actors might face in this context. First-order inference concerns what an actor herself thinks is best; second-order inference concerns what a specific other thinks is best; third-order inference concerns what most others think is best—the conventional belief (Correll et al. 2017; Ridgeway and Cornell 2006). Since Keynes's (1936) famous analogy of the beauty contest, financial markets have been treated as a paradigmatic case of third-order inference. Because an investor profits if they hold stock that others are about to demand, and every investor knows this, each tries to anticipate what most others will do. This is the reason investors monitor prices, financial media, and trading platforms to track where average opinion is heading (Beunza and Stark 2012).

An organization that wants to shape such a valuation process must do more than persuade a particular audience. It needs to *synchronize* audiences' expectations about what the others will do. Abreu and Brunnermeier (2003) showed how public signals can help coordinate the timing of arbitrage by resolving uncertainty about *when* others will act. From a more general perspective,

Schelling (1960) called such events 'focal points:' reference points prominent enough that each actor can assume the others recognize them and can assume the others assume the same. Two people who must meet in New York without agreeing on a place go to Grand Central because each knows that is where the other will expect to be found.

There are a variety of social institutions that can serve as focal points and thereby resolve the coordination problem. Status orderings, awards, rankings, and ratings align divergent expectations in this way because they establish an independent and shared signal about what baseline everyone accepts (Espeland and Sauder 2007). So do public rituals like royal progresses, revolutionary festivals, and circular assemblies (Chwe 2003). Credit ratings are the paradigmatic case: because mandates and regulation tie many investors to the rating, a downgrade tells each holder not only what to think but what every other holder must do (Carruthers 2013; Rona-Tas and Hiss 2010). By stabilizing a common reference point, these kinds of institutions reduce the need to reason from scratch about the likely beliefs of others. In Correll et al.'s (2017) terms, such orderings make conventional belief easier to anticipate even when private preferences differ.

When such institutionalized solutions are absent, actors can sometimes synchronize expectations interactionally. Hollywood writers rework a story mid-pitch while reading the producers' reactions, and comedians anneal the expectations of several audiences in live performance (Elsbach and Kramer 2003; Reilly 2026). In each case, the actors are reading their audiences' reactions and adjusting the performance until expectations settle. However, such interactional techniques only work in contexts where actors are co-present, such that they can read and adjust to their audiences in real time.

This leaves acts of strategic communication where an organization without institutional standing communicates to its audiences via a public text. Such acts of strategic communication

advance a larger claim through some mixture of argument and narrative. While arguments support a claim by presenting evidence and reasons in a logical sequence, a narrative connects events and protagonists into a causal sequence to make a point (Bruner 1991). Examples of such public texts are press releases, earnings statements, or investor prospectuses (Turner 2025). How, then, should the organization compose such a text to crystallize a focal point?

The Baseline Expectation: Focus and Genre Conformity

On Schelling's (1960) definition, a focal point is a prominent signal, whose prominence is itself common knowledge. When organizations hope to produce such a focal point, they should therefore aim to write for broad diffusion. In addition, they should try to ensure that every audience can infer that others will also see the text as credible.

Theory suggests that texts have the best chance to diffuse widely if they are *focused*. In most markets, attention is scarce and the mode in which information is presented can be crucial for whether it spreads (Huberman and Regev 2001). Narratives, in particular, travel because they compress scattered information into a causal sequence that is easy to grasp, remember, and pass on (Beckert 2009; Shiller 2019). The more focused the narrative, the better. Narrative elaboration does not make stories more convincing or easier to understand because audiences tend to complete the suggested causal pattern on their own (Charles and Kendall 2024). More complex accounts also run the risk of being seen as incoherent or poorly supported. Audiences tend to judge a composite object as a whole: a poorly supported part drags down the strong parts around it (Stroube, Vakili, and Bikard 2025). Even audiences who reward unconventional objects penalize incoherent elaborations (Gouvard and Durand 2023). Stories with rich plots are therefore retold less, not more (Heath et al. 2001). Parallel findings exist for arguments: the more focused, the easier the argument is to remember and pass on and the harder to contest (Bennett and Feldman

1981). A text that is meant to diffuse widely should therefore concentrate its material into one line of attack, developed by one dominant storyline or one tight argument.

To become a focal point, a text does not just need to diffuse. It must also be seen as something that others take to be credible. Conforming to an established genre can help because genre conventions signal what counts as credible to everyone. Valuation research has long shown that firms which fit no recognized category may be seen as illegitimate (Hsu 2006; Zuckerman 1999), and may become liable to an atypicality penalty (Cutolo and Ferriani 2024). By analogy, a report that fits no recognizable communicative genre may be discounted as unprofessional and non-credible - particularly when the organizations do not have special standing or operate in settings where there is no preference for stylistic novelty or atypical performances (Betancourt, Hoever, and Wezel 2025). Again, this argument has two dimensions. On the one hand, departures from the generally established standard may simply be read as incompetence (Durand and Khaire 2016; Meléndez, Wood, and Navis 2025). On the other hand, such departures make the coordination problem worse. Zuckerman (2017) argues that the categorical imperative should hold more tightly where evaluators are interdependent because conventions that are common knowledge make coordination smoother than exceptions to them. A report written in the recognized genre of financial research does more than signal competence. It lets each reader assume that every other reader will process the document through the same conventions.

Taken together, this research suggests that simple stories, told in a recognizable genre, should enable the attention, transmission, and credibility that are necessary to create a focal point. However, this line of reasoning imports an assumption from the original definition of focal points. Schelling noted that focal points work only when everyone agrees about the significance of the signal itself, that is, when evaluative standards are settled. In a similar vein, Abreu and

Brunnermeier (2003) saw synchronization as the main problem of arbitrage: investors agree that an asset is overvalued and only need to know *when* others plan to react.

Valuation processes, however, rarely satisfy this assumption. As outlined in the first section, audiences often hold heterogeneous evaluative standards, and an act of strategic communication that seeks to prompt a critical mass of actors must somehow address them. It is not clear what the best way to do so is: a focused narrative may be ambiguous enough to resonate across multiple evaluative standards, but it may also be read as superficial (Paolella and Durand 2016; Pontikes 2012). Conversely, Botelho (2018) found that short sellers share more of their analysis when buy-in uncertainty is high, because their bet pays only if others converge in time. His account explains when short sellers disclose and how much. But it does not address how the disclosure is structured - how the material is organized into the narratives and arguments that heterogeneous audiences read. In short, audience heterogeneity may influence the dynamics of valuation processes where collective judgments matter for profits, but if and how they affect the calculus behind strategic communication is another matter. It is this question we will now investigate.

DATA AND METHODS

Empirical Context: Activist Short Selling in Financial Markets

Short sellers make profits by taking contrarian positions on target companies, betting that the company is overvalued (Botelho 2018; Zuckerman 2012b). The risk of this maneuver can be high: whereas other investors' losses are bounded, short sellers' can be theoretically unlimited, and if the price rises after they publish they may be forced to 'cover' at a loss before their information has been absorbed into the price. Even if they are correct, they may face financial ruin if the market does not move fast enough or turns against them, as in relatively common short squeezes or 'buy the dip' strategies (Stice-Lawrence, Wong, and Zhao 2024). Activist short

sellers therefore publish reports that defend their view, based on original research that includes the analysis of public records (e.g., pitch decks, IPO Prospectuses, earnings calls) and private data (e.g., expert consultations, interviews, site visits). The goal of these reports is to trigger the change in valuation the short seller has bet on. The reports thus represent a form of strategic communication designed to affect the prevailing valuation quickly and dramatically. As an extreme case, these reports are a strategic site to study how organizations use narrative to create a focal point that can coordinate interdependent audiences.

--- Table 1 about here ---

Given the risks of activist short selling, it is a small field. As table 1 shows, most activists in our dataset are generalists who target firms in different industries. While their reports vary in ways that are the topic of this paper, they also follow conventions recognizable as a genre (Paugam et al. 2021; Stolowy, Paugam, and Gendron 2022), illustrated by the two examples in Figure 1. They typically open with bullet-point summaries or an executive overview, followed by thematic sections supported by links, exhibits, and footnotes; many specify a target price for the stock and disclose a short position. While the reports often stand on their own, many are embedded in broader campaign activity that can include social-media amplification, media outreach, and regulatory filings.

--- Figure 1 about here ----

Data Collection

We collected four sources of data: a primary corpus of activist short seller reports; semi-structured interviews with activist short sellers and financial journalists; market data on the target companies; and social media, sell-side-analyst, and business-press coverage. The reports are our primary evidence. We use them to map the rhetorical elements of reports and identify recurrent

strategies. The remaining sources allow us to relate these strategies to characteristics of the target companies and their evaluative environment, which helps to recover the strategic calculus underlying them. Table 2 summarizes the data sources, the variables used, and their distribution and completeness.

--- Table 2 about here ---

Activist short seller reports (short reports). Our analysis draws on a corpus of 1,319 short seller campaigns compiled by a data vendor. We analyze the 873 campaigns targeting U.S.-listed firms, announced between August 2008 and January 2024 (97% from 2015 and onward), of which 851 pressed their case in a written report or substantive multi-tweet thread. We restrict analysis to the 830 texts we recovered in full through automated scraping, web-archive recovery, and manual search. The median report is 4,156 words long, and runs for 19 pages. It contains 16 segments, where each segment is one coherent argument or narrative, and 7 exhibits.

Interviews. Between August 2024 and July 2026 we conducted 21 semi-structured interviews, 16 with short sellers and 5 with financial journalists who cover the field, including three who moonlit as analysts for short sellers. Interviews averaged sixty-five minutes and were transcribed for analysis. They give us access to short-sellers' own accounts of investigative material collected, allegations to make, audiences to target, and expected reactions. To preserve anonymity in this small field, we use pseudonyms and general descriptions of role and experience. Appendix B describes the interview sample.

Audience data. For cases studied in depth, we captured reactions from major audiences: sell-side reports from Refinitiv and LSEG Aftermarket Research, earnings-call transcripts from Capital IQ, and business-press coverage. For retail investors, we collected Reddit posts referencing the target ticker in a ±100-day window around the campaign, along with commentary on X

and Seeking Alpha.

Market data. To characterize the evaluative context, company characteristics, and report impact, we collected market and accounting data. From CRSP daily data we compute market-model cumulative abnormal returns (CARs) and target size; from Compustat quarterly filings available before the announcement, we take industry affiliation, leverage, asset growth (COOPER, GULEN, and SCHILL 2008), and net operating assets (Hirshleifer et al. 2004). These variables control for substantive reasons a company may be targeted, over and beyond the presence or absence of stable evaluative standards.

We measure the presence of shared evaluative standards with an adaptation of what Botelho (2018) referred to as 'buy-in' uncertainty. The measure averages five components — firm age, analyst coverage, media coverage, and two facets of institutional ownership — each proxying an evaluative institution that monitors the firm; denser evaluator infrastructure means lower uncertainty. The measure is structural: it registers whether the institutional infrastructure for evaluating a firm is in place. The underlying intuition is that older firms with high institutional ownership and broad analyst coverage are generally better understood by the market, and thus subject to more stable evaluative standards. We call the measure 'evaluative' rather than 'buy-in' uncertainty because it captures more broadly whether there is an established way to think about the company, rather than whether someone will buy or sell its stock.

Even when evaluative standards are present, these may not translate into stable valuation. A firm can be subject to a clear standard yet trade at a relatively uncontested price – and vice versa. To distinguish the presence of evaluative standards from active agreement, we compute the target's realized volatility over the 60 trading days before the campaign. While low volatility indicates relatively high agreement on the price, high volatility is a trace of investors persistently

revising against one another's expectations. The two measures are related but distinct (Spearman $\rho = 0.4$); Appendix A provides further details about the construction of the different measures. It shows that the volatility measure is correlated with a direct measure of analyst dispersion, which measures disagreement directly but only for a smaller sample.

Data Analysis

We analyzed these data in four stages that correspond to the sequence of findings. First, we combined close reading and corpus-scale coding to map variation in rhetorical architecture. Second, we analyzed interviews to recover the coordination problem as practitioners understood it. Third, we related the corpus-level configurations to evaluative uncertainty and active contestation while probing alternative explanations. Fourth, we returned to selected cases to interpret the strategic logic and refine the typology. The process was abductive rather than linear: later stages prompted us to revisit earlier distinctions.

Step 1: Mapping Rhetorical Architecture. We developed an inductive codebook by closely reading reports across short sellers, target industries, and time periods. For an initial round of qualitative analysis, we randomly sampled and coded reports to identify the core rhetorical elements of these reports. Both authors coded in parallel, collectively wrote approximately 40 report memos and 20 theory memos, and met for integrative discussions. Our first-level inductive codes included narrative elements (e.g., character, central action, story tropes), style of argument, including logico-scientific passages (e.g., marshaled evidence, scientific reasoning) and rhetorical strategies (e.g., accusing, casting suspicion, appeals to credibility, humor and irony). After reading approximately 120 reports, we converged on higher-order categories that capture how short sellers write and what stories they tell (Gioia, Corley, and Hamilton 2013; Grodal, Anteby, and Holm 2021).

Specifically, we identified three features that characterize a passage's rhetorical structure: its tone, its use of narrative form, and the substantive stories it tells. Tone runs from professional through colorful to theatrical. It captures the extent to which the author performs for the reader with hyperbole, irony, and sarcasm, rather than merely informing. For narrative form, we register whether a passage uses narrative form to deliver a claim or present a pure argument. A narrative is defined as a sequence of at least two causally connected events, while an argument is a claim that is backed by some warrant. We also code for nine distinct tropes — substantive story arcs that explain why a company is overvalued, such as shaky foundations, boom-bust, hidden skeletons, or hustler stories. Together, these three features describe *how* a passage advances a given line of argument. Table 3 contains definitions and examples of these higher-order codes.[2]

--- Table 3 about here ---

Corpus-scale coding. Once close reading stopped producing new categories, we froze the codebook. While this departs from the prescriptions of grounded theory (Charmaz 2006), it is necessary to apply the codebook reliably to the whole corpus (Nelson 2020); we continue in the spirit of grounded theory on a different level, by using the patterns revealed at the corpus level to inform further in-depth case studies that allow us to theorize these patterns as distinct strategies.

We then engaged LLM agents to code the whole corpus, following the approach of 'computational grounded theory' (Nelson 2020; Than et al. 2025) with qualifications outlined in Appendix E. We used an ensemble of three frontier models (Claude Sonnet 4.6, GPT-5.1, GPT-5-mini) to code each segment of the corpus under structured prompts based on the same manual

[2] The full codebook with additional categories, particularly a count of ten substantive allegations, is available upon request.

human coders used. Because segment-level agreement was relatively *low* even under a strict majority vote, we aggregated segment-level votes to campaign-level constructs. Table 4 reports the reliability metrics for the aggregate constructs used in the analysis. Appendix E provides further details on how we built and validated the campaign-level metrics.

--- Table 4 about here ---

Step 2: Recovering the Coordination Problem. We analyzed the interview transcripts iteratively for statements about the report's intended audiences, the order in which different actors encountered and reacted to it, and what each actor was expected to infer about later audiences. We compared these accounts across interviewees, using points of agreement and disagreement to refine the sequence. This analysis shifted our interpretation from generalized synchronization to a sequential coordination problem based on second-order inference. The interviews also repeatedly pointed to two difficulties—audiences could lack a stable evaluative standard, and rival valuations could already be publicly advanced and challenged—but at this stage we did not yet treat them as distinct environmental conditions.

Step 3: Relating Architecture to the Evaluative Environment. We then returned to the corpus to assess whether the architectural patterns varied with the two difficulties suggested by the interviews. K-means clustering of the campaign-level trope codes identified a two-cluster solution for rhetorical breadth — the most balanced and stable under resampling among the alternatives considered (fit statistics in Appendix C). We treated narrative share as a continuous corpus-level indicator of narrative integration and tone as a descriptive aggregate rather than clustering either measure. We examined the associations of rhetorical breadth and narrative share with evaluative uncertainty and pre-campaign volatility, and considered whether the patterns instead reflected seller-specific house style, report length, allegation count, target-business complexity,

or the supply of unexplored material.

To estimate these relationships, we fit campaign-level OLS models with standard errors clustered by short seller. The first specification is unadjusted. The core specification adds announcement-year and two-digit NAICS fixed effects and book leverage; the full specification additionally includes asset growth and net operating assets, measured from the latest pre-announcement filing. We do not include firm age or market capitalization in models of evaluative uncertainty. Age is an EUS component, and size closely proxies the evaluative infrastructure the index is intended to capture; adjusting for either would partial out part of the outcome itself. We apply the same control ladder to the volatility models so that the breadth–uncertainty and breadth–volatility estimates form a common discriminant comparison. Separate robustness analyses examine report length, visual-exhibit depth, allegation domains, target complexity, and short-seller fixed effects. Comparing these patterns separated evaluative uncertainty from active contestation. The analyses established contingent associations, not intent or causal effects; they defined the configurations and informed the comparative-case sampling in Step 4.

Step 4: Interpreting Strategies Through Comparative Cases. To understand the strategic logic behind the corpus-level associations, we returned to in-depth case studies. We selected 64 campaigns from a condition matrix crossing high and low evaluative uncertainty with high and low pre-campaign volatility, our indicator of active contestation, sampling typical, deviant, and boundary configurations. For each case, we analyzed the report alongside interviews and reactions from news media, target companies, retail investors, and institutional investors. We used this comparison to determine whether broad lines remained modular or were integrated into a report-level throughline, to interpret how each configuration addressed the coordination problem, and to identify its vulnerabilities. Market outcomes diversified the sample but did not define

the configurations; reception evidence for atypical cases revealed that the drawbacks the literature has identified for broad narrative and unconventional presentation are real. Adding cases until theoretical saturation produced the typology of focused, modular, and integrated strategies.

Although we began inductively, the back-and-forth between corpus, interviews, and cases shaped the analysis into a computationally supported abductive process, moving between empirical observations and existing theory (Tavory and Timmermans 2014; Timmermans and Tavory 2022).[3] The focus on evaluative uncertainty, for instance, emerged when we noticed how often campaigns referenced other audiences ('insiders are getting out') and asked interviewees about it. Similarly, the difference between evaluative uncertainty and volatility, which we had treated as interchangeable, first emerged in the associational analysis. The qualitative and computational approaches thus disciplined each other in an abductive process that gradually led to our theory.

FINDINGS

We develop our findings in four steps. First, we characterize variation in the rhetorical architecture of short-seller reports. We show that reports exhibit recognizable house styles, but that short sellers also adjust two rhetorical dimensions of their reports to target companies: the rhetorical breadth of their reports; and the degree to which they integrate the overall report with narrative. To understand what drives these decisions, we then turn to interviews and show that short sellers understand themselves to be solving a coordination problem different from the one typically assumed. Rather than facilitating generalized third-order inference among market audiences, they try to initiate an anticipated cascade in which earlier actors infer the likely behavior

[3] We found one precedent of abductive computational theorizing that deploys the insights from computational analysis in a similar way, though they use different content analytical tools (Karell and Freedman 2019).

of specific later audiences. The interviews suggest that evaluative uncertainty and market volatility make this anticipatory task difficult in different ways. Third, we show that rhetorical breadth is associated with evaluative uncertainty. Within broad reports, narrative integration is associated with volatility. We return to detailed case studies of typical cases to recover the strategic logic behind the three archetypical configurations: focused, modular, and integrated reports.

Variation in Rhetorical Architecture of Short Seller Campaigns

Every report develops the same overarching claim that the target company is overvalued. Nonetheless, reports vary in how they select and organize the material used to advance that claim. We distinguish two core dimensions of reports' rhetorical architecture. The rhetorical breadth captures the number and range of substantive narratives that a report develops. Narrative *integration* captures the extent to which a report unifies these lines of attack into a single narrative throughline or keeps them separate. *Tone* describes the use of stylistic devices and ranges from professional to theatrical. We discuss these dimensions in turn.

Every report is a mix of stories, arguments, and exhibits. We identified nine recurrent stories, or tropes, that short-sellers tell to enliven their arguments. These are detailed in Table 3 and can be broadly grouped into stories about people, about the company, and the market. The common 'hustler' trope, for instance, is a story about someone at the company engaging in fraudulent activity. Conversely, a 'shaky foundations' or 'impossible challenge' story explains why the company is either built on a flawed business model or confronts an obstacle it cannot reasonably overcome. A boom-bust story, finally, suggests that the company's valuation is affected by inflated values for the entire industry. Together with substantive arguments and analyses, these narratives define the various lines of attack that constitute the substantive case the report advances against the target company.

A first question was how short sellers combine these different stories in their reports. K-means clustering reveals two basic configurations. They are illustrated by Figure 2, which shows how the two clusters deviate from the corpus average in their use of the nine tropes. The smaller cluster contains *focused* campaigns that anchor in one dominant storyline, such as corporate obfuscation, fraud, or reckless stock promotion. In focused reports, only three or four storylines appear at all (3.7 on average) and the lead story dominates nearly half of the report's narrative passages. By contrast, *broad-repertoire* campaigns run nearly twice as many storylines (7.1) with no clear lead: stories about hidden skeletons sit beside stories about compromised CEOs, and doomed business models. Because most reports belong to the broad cluster, short sellers often depart from the theoretical expectation of focus by developing multiple distinct storylines.

----- Figure 2 about here -----

Apart from rhetorical breadth, reports also vary in *how* their claims are advanced. A substantive passage may be presented in narrative or in argumentative form. Whereas a narrative links at least two events in a causal sequence to make a point, an argument asserts a claim that is linked more or less explicitly to some warrant. For example, "The company tried to pivot to SaaS, but this ultimately led to margin compression" is a narrative. By contrast, "the company faces three problems: margins below peers, debt too high, and no insider buying" is an argument. We use the percentage of passages that use the narrative form as a measure of *narrative integration*. Figure 3 represents the distribution of this variable.

--- Figure 3 about here ---

The upper panel shows that narrative share, our indicator of narrative integration, is concentrated toward the high end but has a substantial lower tail. The average report relies on narrative for over 80% of its segments, but the interquartile range extends from 74% to 94%. In

depth-reading of cases with high levels of narrative (more than 90%) average levels (80%), and low levels (less than 50%) suggests that high levels capture the presence of an overarching story that pulls the different pieces of the report together. Conversely, more moderate levels of narrative point to reports that mix neutral arguments with evocative stories and low levels of narrative refer to reports that mostly discuss evidence.

Tone captures whether the author presents the material professionally or relies on colorful or theatrical devices such as hyperbole, rhetorical questions, or sarcasm. The distribution in the upper panel of Figure 3 shows that average tone is concentrated near *professional* but has a meaningful theatrical tail. Very few campaigns are written entirely in the professional language of financial analysis. Most punctuate neutral accounts with colorful and theatrical passages, often concentrated in the front-end, while others rely on hyperbole, sarcasm and internet memes throughout.

Reports thus deliver their content in different tonal registers, different mixes of arguments and narratives, and different levels of rhetorical breadth. For illustration, consider the two examples in Figure 1. White Diamond's attack on CleanSpark is short and punchy but develops multiple narratives about the company; Spruce Point's report on AECOM is long and sober but develops a focused accounting story next to a variety of arguments. Along these dimensions, reports frequently depart from the baseline expectation of the theory. High levels of narrative integration and colorful tone deviate from the conventions of financial research while broad repertoires violate the expectation that activist short sellers should aim to be focused and concise.

What, then, explains this variation in tone, narrative form, and content? In our interviews and in the literature on short-seller rhetoric (Stolowy et al. 2022), one answer dominates: brand. Sellers cultivate a recognizable voice because reputation is itself a coordinating signal – and once

a brand is established, it carries its own signal of legitimacy. One interviewee recalled the influence of a legendary activist: “there was a period of time when David [Einhorn] could send stocks down by just asking a question at a conference call.” Indeed, sellers do have clear rhetorical signatures. Table 5 decomposes how much of the variation in a report can be attributed to a fixed house style and how much lies within a seller’s own catalog.

--- Table 5 about here ---

Tone is the clearest signature, with half its variance between sellers ($\eta^2 = 0.50$): Spruce Point writes almost everything in the flat register, while Citron, the polar opposite, tends to rely on more colorful language. The choice between broad and focused repertoires is much less seller specific ($\eta^2 = 0.30$) as is the reliance on narrative versus argument ($\eta^2 = 0.18$), but a careful comparison of reports from the same short seller supports the claims from interviews. For instance, Spruce Point is often very close to a standard institutional analysis. Apart from using professional language, it typically develops its line of critique with arguments that are supported by extensive exhibits. By contrast, the Friendly Bear writes in a much more colorful tone that combines memes, internet-humor, and other rhetorical devices with vivid narratives that heap ridicule on the companies they target. For instance, its attack on Freshpet was titled “Combine A Bulldog With A Shih Tzu, And You Get Freshpet” and opens not with the target but with a joke at its boosters’ expense — announcing a mock "All-Hype research award" for sell-side analysts who prop up absurd companies and naming Freshpet’s own underwriters as the first nominees.

The effort to develop a recognizable house-style offers a partial explanation for the deviation from the genre of financial investment research. If the brand is recognized as credible, the house style performs the same signaling function as the genre conventions otherwise do. Nonetheless, *most* of the variation lies within sellers’ catalogs: the same activist writes focused and

broad reports, as well as more or less integrated reports depending on the target. House style therefore sets a recognizable baseline, especially for tone, but does not exhaust report-level variation. This leaves the target-level question open: why do short sellers vary the rhetorical architecture of their reports relative to the targets they attack? If tone is largely house style, what drives the choice between focused and broad content on the one hand, and a more or less narrative integration on the other?

The Coordination Problem Short Sellers Try to Solve

When we started to explore why short sellers varied the rhetorical architecture of their reports relative to target companies, we assumed that they were trying to solve the collective action problem posited by the literature: synchronizing generalized market expectations. But the interviews forced us to revise that interpretation. Activists described a sequential process in which investors try to understand what specific later audiences are thinking. According to our interviewees, a report had to serve as a "catalyst" that triggered an attention cascade from one audience to the next. Most interviewees agreed that the primary target was investors who already owned the stock or were likely to buy it, a heterogeneous group ranging from relatively naïve retail investors to institutional players such as hedge funds, family offices, and endowments. Yet these actors were not the first to react. They were later points in a cascade that begins with algorithms and speculative actors.

When a report hits "the Bloomberg terminal," algorithms determine where it appears in traders' feeds. "Your first reader is going to be a machine," said Charlie. Quinn, who pushed the publish button at his firm, recalled prices reacting "sub-one second, sometimes on a pretty signif-

icant scale" as quantitative funds "pinged the site constantly" and parsing the report with "language processors." These algorithms were looking for trigger words that signaled likely relevance for human traders. Some short sellers proposed theories about how to trigger them. For example, Winter said, "If you use the word fraud… it triggers all the algos." In any case, due to the automated trading, a report's success is "usually defined in the first few minutes of the trade," said Riley.

But once algorithms have picked up and amplified the report, "there's always an initial candle. I think it's people who scrape [our website for information]." The first human actors are thus what Thomas referred to as "scalpers… who look at people scraping, looking for a quick trade." Then it's the "options market-makers," whose hedging "brings in the momentum." At that point, retail investors pile in: "They're playing for a couple percent. They're not playing for a lot," said Thomas, followed by low-conviction momentum holders who sell or, in Sam's words, "just … puke it." Until this point very few people, if any, will have read beyond the headline and introduction of the report. Speculative actors seek quick signals that help them to anticipate the reactions of specific market actors, i.e. who will sell, when, and in response to what kind of cue.

This dynamic shifts in the next phase: Only "after all those guys have come to the party, it's when we find out whether it's going to stick," explained Thomas, as the long institutions "finally get around to reading it after it's down 10, 15%." At that point, the broader interpretative field — media, analysts, the target itself — begins to respond. These actors aim to assess the substantive plausibility of the short thesis: its evidentiary and financial credibility. They are likely to read the analysis in full or in part, and may replicate it using their own information sources. But they also look at what other people are doing. As Saul remarked: "Everybody looks

at what other people are doing. You just do. It's human nature."

The short seller thus has to trigger an initial cascade of algorithmic and momentum-based trading that then generates attention among long investors who hold substantial positions in the stock. Short sellers noted that the task to get the cascade going could be more or less difficult. As AD. explained, short sellers were "really trying to compete for very limited attention spans" and therefore had to "distill the argument into … the one or two best things." Charlie described contemporary activism as "a headline driven game" in which sellers "gamify" their presentation to produce the "bombastic blockbuster headlines" that trigger algorithms. This pressure was particularly acute when a given stock already experienced contestation, when it was already "in play." Describing the volatile, retail-dominated stocks he targeted, one short seller drew the connection directly: "they're always in play, so they have so many different participants […] I think the main thing is, like, how can you capture an audience? Everyone these days has low attention spans — especially the stocks that I'm going for, the more retail names. […] You post a report more than two pages, no one's going to the third page." Martes similarly suggested that for highly contested stocks you have to ask yourself "what are the bullet points that will float on Twitter?"

But apart from successfully capturing attention and diffusing their report in the early stages of the cascade, short-sellers also needed to make sure that they enable audiences to infer what everyone along the chain is going to think. While algorithms look for mechanical signals that will likely trigger broader market reaction, the early human traders are trying to assess how the specific actors who appear later in the sequence are likely to react. The long-investors who begin to act when the market has already moved, try to decipher the intentions of other long investors who evaluate the report on its merits. This sequential structure implies that investors are

not only trying to anticipate the conventional belief — what most others think is best. They must also determine how the specific audiences that act after them will react. They face not just a problem of third-order but of second-order inference.

This means that short sellers must be careful to understand the evaluative standards of the different audiences along the cascade. This, according to Zeke, affected not the "research process [but] the story structuring." Depending on the target company, audiences might be more or less aligned on these standards. Of course, all short sellers agreed that audiences were looking for information that was "material" and "provable" by investigators, but they noted frequent disagreement among audiences about what information met these standards. As Arthur explained, for a "business that's been around for 100 years" the standard is settled: "you're gonna talk about kind of GAAP earnings… because that's how those businesses are valued." In contrast, audiences for tech companies often focus on other indicators, such as revenue or user growth. Evaluative uncertainty is especially pronounced for younger firms offering novel and unproven technologies, where investors may abandon conventional standards without agreeing on what should replace them. As Arthur put it: "It's a lot easier for people that are selling that new technology to investors to convince them that old valuation metrics no longer are reasonable. [But then] it's difficult to value it." In that case, audiences might vary substantially in how they evaluated new information. Marion described a rockets company whose retail audience valued "when the rocket was launching," while institutions saw it as a defense contractor. Short-sellers thus highlighted that target companies varied in the extent to which audiences shared a stable evaluative baseline – the degree of evaluative uncertainty.

Three Rhetorical Strategies: Focused, Modular, and Integrated Reports

Taken together, short sellers perceived two primary challenges in triggering the desired

cascade: securing attention among algorithms and speculative actors during its early stages, and enabling audiences not only to evaluate the thesis but also to infer the likely views of other audiences along the chain. Interviewees suggested that these tasks become difficult in different ways depending on the target's evaluative environment. Evaluative uncertainty captures whether audiences share a stable standard. Active contestation captures whether rival valuations are already publicly advanced and challenged. Table 6 reports campaign-level associations between these conditions and the report's rhetorical architecture. Because active contestation is not directly observable at scale, pre-campaign volatility serves as its empirical indicator.

---- Table 6 about here ---

Rhetorical breadth is greater around firms with more evaluative uncertainty: the raw broad–focused difference is .37 standard deviations (Welch $t = 5.08$, $p < .001$), .32 SD after adding leverage and announcement-year and industry fixed effects, and .29 SD after additionally controlling asset growth and net operating assets. Standard errors are clustered by short seller. Pre-campaign volatility does not distinguish broad from focused reports. Instead, narrative share, our measure of narrative integration, rises with volatility within broad reports but is approximately unrelated to volatility within focused reports. Regressing narrative share on the interaction of report breadth and volatility yields a slope difference of about +.04 per standard deviation of volatility ($p \leq .05$ across the full control ladder). This estimate is sharpened by short-seller fixed effects (+.04 to +.05, $p \leq .02$). The same seller relies more heavily on narrative when a broad report targets a more volatile firm. When substituting volatility with analyst forecast dispersion, which is a more direct measure of disagreement, this this interaction reappears on the covered subsample (+.03, $p = .04$ with controls; +.04, $p = .04$ with seller fixed effects, see Appendix D). This is the case, even though dispersion is defined only for larger, analyst-covered

targets, so this check runs on a calmer, better-covered segment of the corpus. These findings suggest that evaluative uncertainty maps onto rhetorical breadth, whereas active contestation maps onto narrative integration among broad reports. Figure 5 displays the resulting configurations.

--- Figure 5 about here ---

Before interpreting these configurations as strategies, we considered several more mechanical explanations. Broader repertoires might simply reflect longer reports with more information, more numerous allegations, more complex target businesses, or stable seller-specific styles. Supplementary analyses show that the breadth–uncertainty association survives controls for report word count and visual-exhibit depth, the number and mix of ten allegation domains, year and industry, and a specification with short-seller fixed effects. Nor is breadth explained by target complexity, measured using business- and geographic-segment counts, segment-sales concentration, intangible-asset and goodwill shares, and the length and complexity-word share of the target's most recent pre-campaign 10-K. No individual complexity measure correlates with breadth above $r = 0.09$ (Appendix D, especially Table D2). Broader reports are therefore not merely longer dossiers or mechanical responses to more allegations, more complex firms, particular industries, or particular sellers.

The seller-fixed-effects specification carries substantive weight beyond robustness: the same seller is more likely to write a broad report when facing a higher-uncertainty target. The within-seller probability of a broad repertoire rises by about .07 per standard deviation of evaluative uncertainty, $p < .001$; Table D2, final row. This is another indicator that we are tracking a rhetorical strategy that cannot be reduced to the creation of a house style.

Taken together with the conditional narrative-share result above, the evidence suggests

that reports are organized differently depending on the evaluative uncertainty and volatility surrounding the target. These associations identify a contingent pattern in reports' rhetorical architecture, but they do not by themselves establish its strategic logic. We therefore draw on interviews and close readings to examine how each report type addresses the coordination problem identified above.

1. Low Uncertainty: Focused Reports

Focused reports tend to target companies that audiences evaluate against settled evaluative standards, often older companies with a proven business model and track record, for which clear standards allow comparative judgments about performance. We identify two types of focused reports: The first type is concise and anchored to an accepted standard of evaluation with a small number of well-substantiated claims. A good example is Citron's report against TransDigm Group Inc. Written in Citron's flamboyant house style, the ten-page report attacks a large, well-followed aerospace company, whose numbers are public and stable. The core argument is narrow. It relies on standard measures of business analysis (peer EV/EBITDA multiples, gross-margin comparisons against peers) to suggest that the company masks negative organic growth close to 10% with repeated price hikes — hikes long accepted by the government, but that the incoming Trump government had just vowed to end. The report centrally relies on the hustler story trope, ending with a direct appeal to Trump himself:

> "President Trump: Do you know when your 757 plane "Air Trump" is down for a day and you wondered why a replacement air filter, valve, pump or coffee maker cost $20,000? The reason was TransDigm. While you are looking to Boeing and Lockheed, the easiest way to drain the swamp in the aerospace industry can be relegated to shine some sunlight on: TransDigm"

The focused strategy typically develops a single line of argument that shows how new information shifts a received judgment based on a settled standard. In this case, Citron shows that the profitability of the company hinges on a practice that faces a risk of being curtailed.

The second type of a focused strategy aims at reclassification, arguing that valuation should pivot to an alternative, but similarly accepted standard. One example is the Friendly Bear's report against Applied Digital. Written in the muckraker style, the report argues that the AI-datacenter company should be evaluated from the standpoint of litigation risk rather than business fundamentals, focusing entirely on how its governance structure creates potentially illegal relationships to its patron investment bank, B. Riley: "Applied Digital represents one of the worst governance situations we have ever seen." Though the report works almost entirely in a legal-forensic register, it combines the technical material with hustler narratives that show how compromised board members hold the company captive, sharpened by rhetorical flourishes ('does not pass the smell test') and sarcastic remarks. At its core, it provides narrative and argumentative evidence for a narrow claim that suggests re-evaluation under a commonly accepted standard.

Focused reports closely follow the theoretical expectation of how focal points should be established. Because they are focused, they can diffuse easily, even when attention is limited. As interviewees affirmed, whenever possible you "want to lead with a story that is just very notorious [and] as obvious as possible," suggested MS. This is possible because the audiences have a shared evaluative standard. At every point in the cascade, investors can assume that other actors will evaluate the thesis on the same standards as them. This shifts attention to the validity of the thesis relative to that standard. The response to Citron's report against TransDigm, for instance,

was focused on the question whether prices were truly inflated. The business press treated the report's timing on inauguration day as opportunistic, but then focused on the credibility of the evidence presented. Presumably because focused reports are already able to diffuse easily, there seems to be no relationship between volatility and their rhetorical architecture. Friendly Bear's attack against the volatile Applied Digital followed the same playbook as Citron's against the uncontested TransDigm. Table 7 lists details for these campaigns and summarizes additional examples that informed our analysis.

--- Table 7 about here ---

2. High Evaluative Uncertainty and Limited Contestation: Modular Reports

Stable evaluative standards are not always available. Where a company is newly listed, diversified across market segments, tightly held, or has recently pivoted, audiences may not yet have settled on a shared way to evaluate it. This evaluative uncertainty does not always produce active contestation. In some cases, audiences have arrived at a relatively stable price from different evaluative directions. Short sellers targeting these companies face fewer competing public accounts but more unsettled standards. In this environment, activists such as Spruce Point Capital, Prescience Point, and Gotham City Research use modular reports: rhetorically broad documents that keep their lines of evaluation separable.

Gotham's sixty-page report against Endurance International Group from 2015 is a characteristic example. Though founded in 1997, the company had been listed for only a year and a half and had not drawn much media or analyst attention yet. The report develops a set of separable arguments, each pitched to a different standard of judgment. It opens by identifying a paradox that a company with "industry-leading margins" is nonetheless "bleeding cash" like one of

the worst companies. It then develops separable arguments that resolve this puzzle from different directions. The first piece is a financial analysis of revenue quality, based on the internal documents of the company, showing that "Average Revenue Per User (ARPU) declined -13%." The argument is supported not just by accounting evidence, but also by narrative episodes that make user dissatisfaction tangible:

> Endurance spends 60-80% less per customer on infrastructure, than Godaddy does. […] Endurance companies [therefore] experienced at least 4 large-scale service outages last year, and they seem to be worsening. […] Shortly before 3 p.m. Tuesday, BlueHost's official Twitter account tweeted the following announcement: "FYI we're experiencing a temporary network issue with our data center that's affecting some customers. If that's you stay tuned for updates!" The outage led a number of BlueHost and HostGator customers to send tweets of frustration to the accounts for the hosting services.

Another line of reasoning focuses on fraud perpetrated through subsidiaries to enrich board members, explicitly likening the company to frauds like Crazy Eddie, the U.S. electronics retailer of the 1980s. Yet another documents that Endurance's brands "host terrorist websites," with rich narratives detailing the reputational and regulatory risks associated with such customers. A fourth suggests that the company is not equipped for a newly competitive environment, because its earnings are falling below the interest it has to pay on its debt. The different lines of argument feed into an overarching thesis — the company is a 'web of deceit' that is going to tumble soon — but they do not logically or narratively depend on each other. Interviewees referred to this style of report as "kitchen-sinking." They advance a variety of different claims without clearly connecting them.

The structure of modular reports suggests that their goal is coordination rather than appealing to the evaluative criteria of any one audience. In studies of category spanning, breadth reduces the appeal of a product because audiences cannot establish which category it fits (Hsu 2006). A modular report, however, does not need to be liked by every audience. Short sellers need *each* audience to see that every other audience's standard has been addressed, so that each

can infer that the others will act. This explains why short sellers do not tailor separate documents to separate audiences, as film studios segment their marketing (Hsu 2006). They need multiple different audiences along the sequence to converge on, and convergence requires that the coverage of divergent standards be visible to all audiences at once. In rituals, common knowledge arises because participants can see one another attending to the same signal (Chwe 2000). A report is read privately, so the text must substitute for co-presence: the visible breadth of its table of contents shows every reader what every other reader has been shown. Breadth is likelier to be read as versatility than as incoherence where audiences evaluate the report against their own goals rather than against a genre prototype (Boulongne and Durand 2021).

This is why modular reports often include lines of argument that are weakly evidenced and give them equal weight in the table of contents. Spruce Point's report on Cintas Corporation, for instance, pairs a well-supported case of fire-inspection fraud with accounts-receivable claims resting almost entirely on anonymous insider interviews. If those lines were meant to convince the audiences who care about receivables, their weak evidence is hard to explain; as our interviews suggest, these passages address readers who skim the opening pages and see that several issues are covered. Keeping the lines separable also contains the cost of this breadth. Audiences tend to judge a composite objects as a whole, so a weak part drags down the strong parts around it — but the penalty attenuates when readers can extract individual claims from the bundle (Stroube et al. 2025). A modular architecture invites exactly this reading: journalists and analysts can pull the best-documented allegation and circulate it on its own.

Modular reports may therefore expand not only to persuade each audience separately but also to make the range of likely audience reactions publicly visible. This is also consistent with interviewees' observation that the length of these kinds of reports was often itself an important

indicator. One short seller recalled publishing a report in 2011. A half-day of trading and media coverage followed, while the stock "tanked," before a reader noticed that, due to a PDF conversion error, only half of the 88-page document had been posted. Particularly early in the cascade, signaled relevance seems to eclipse demonstrated relevance. Table 7 details additional cases.

3. High Evaluative Uncertainty and Active Contestation: Integrated Reports

When evaluative uncertainty is high, volatility is more often high as well, because uncertainty about evaluative standards easily translates into active contestation. Often, these are 'meme-stock' companies, poorly covered new listings, failed-trial biotechs, or companies with bankruptcy-risk. Here, short sellers face a dual challenge: audiences disagree about the correct way to evaluate the company, and the active disagreement surrounding the target makes it difficult to be heard while increasing the risk of short selling. Such volatile targets are disproportionately the province of newer activists, such as Hindenburg, Culper, or Wolfpack. Apart from relying on broad repertoires of tropes and arguments, they *integrate* these accounts with an overarching narrative.

Hindenburg's 2020 report on Nikola is the canonical case in this segment. Nikola was a pre-revenue hydrogen-truck company that had gone public through a SPAC (Special Purpose Acquisition Company) and was valued near $20 billion on the strength of its promises alone. With no product and no revenue there was no settled way to price it. After a headline-grabbing partnership with General Motors was announced, speculative investors had run up Nikola's stock.

Hindenburg disrupts this speculative rally with a report that knits a variety of arguments and narratives together into a single, overarching hustler narrative of the founder as a fraud. It opens with a rhetorical question — 'What is the difference between a visionary selling a daring

view of the future and a con artist?' — and then uses the biography of founder Trevor Milton as a narrative throughline to develop a variety of disparate revelations about the company. Most famously, Hindenburg shows that a promotional video of the EV truck driving 60 mph "was an elaborate ruse — Nikola had the truck towed to the top of a hill [...] and simply filmed it rolling down the hill." Hindenburg recreated the stunt at the same location and added former employees' testimony about the fraud. The report separately shows that the order book was "filled with fluff," and that the CEO appointed his brother to run hydrogen production even though his "prior experience looks to have largely consisted of pouring concrete driveways." Each claim could have been advanced separately, as in a modular report, but Hindenburg did not leave them standing side by side. It threaded them chronologically into a narrative of a man who "lies like most people breathe," ending with fifty-three numbered questions addressed directly to Milton, the last of which asks "Have you ever deceived anyone?" This report, like many others in this segment, relies on a moral narrative to tie the report together.

Another way reports build narrative integration is through a narrative arc that culminates in a coordinating event. Cliffside Research's 2016 report on Clearside Biomedical—a pre-revenue biotech whose stock had tripled since its June IPO—does not tell a story about a villain. Instead, it pulls multiple strands of argument into a timeline that predicts a sell-off. A re-derived patient count, a cautionary drug history, an anatomical limit, and insider red flags form a chronological progression: the company's drug launch will fail, and insiders will exit as soon as they legally can. "12.3 million shares come off lock-up at the end of the month. Buyers here may be left holding the bag." Interviews and case comparisons suggest that narrative integration helps a broad report compete for limited attention. Although Hindenburg's report against Nikola exceeded eighty pages, audiences mainly recalled the story about the fraudster, immortalized in the

image of the truck rolling down the hill. Table 7 details additional cases.

If narrative integration helps broad reports diffuse, why do modular reports exist at all? Interviews and fragmentary audience data suggest a possible answer. Excessive use of narrative and overly evocative language can be seen as manipulative and unprofessional. Veteran activist Jackie suggested: “The more hyperbolic [a report is], typically, I would look even harder to see that there was real evidence being presented.” Similarly, when Aurelius Value and Grizzly Research targeted the payment processor Intelligent Systems in 2019, CEO Leland Strange conceded on the call that the facts were not in dispute. His objection was that the reports “stream them together in a misleading narrative.” While narrative integration can improve the diffusion of a report, it can also expose the communicator to accusations that selection and causal ordering manipulate the evidence. Modular reports allow individual claims to stand on their own and therefore mitigate this risk.

A final piece of evidence corroborates the theory that rhetorical fit has strategic value. In a supplementary analysis, we compare the cumulative abnormal returns of campaigns that fit their evaluative environment with those that do not. For the content dimension, fit either means that a broad repertoire is used for a top-tertile-uncertainty target, or that a focused one is used against a bottom-tertile target. For narrative integration, a top-tertile narrative share is used on a top-tertile-volatility target or bottom-tertile narrative share is used on a bottom-tertile volatility target. The middle ranges are deliberately excluded. As Table D3 shows, campaigns whose rhetorical strategy matches the target environment on both dimensions show a larger abnormal decline in the month after the campaign. The joint-fit coefficient is +.089 in the 21 day window after the campaign ($p = 0.037$, $N = 351$, standard errors clustered by seller). This association is stable across firm and year controls and short-seller fixed effects. Consistent with the idea that fit

matters for coordinating perceptions along a cascade of different audiences, the effect is absent for shorter time window that capture the announcement shock of a campaign. Because the sample with valid abnormal returns over-represents covered, lower-uncertainty targets, the evidence should be read as corroboration of the strategic logic, not a test of effectiveness.

DISCUSSION & CONCLUSION

Summary of Findings

Short sellers try to create a focal point that allows actors to infer how others will react to the same information. Our findings suggest that the evaluative environment of the target company changes the nature of this coordination problem, and that short sellers configure the report's rhetorical architecture in response. When evaluative standards are settled, every audience judges by the same standard, so the problem of anticipating specific others reduces to the familiar problem of anticipating the conventional belief. Short sellers accordingly write reports that follow the standard prescriptions: focused and conventional accounts are built to diffuse easily and to signal credibility. Because such a report defends its claim against a shared standard, each investor can infer what others are likely to believe about its substance. When evaluative standards are unsettled, third-order inference becomes more difficult and turns into second-order inference: audiences need to infer the likely reactions of specific others who may judge the report differently. In modular reports, short sellers use rhetorical breadth to make visible that several standards, and the audiences attached to them, have been addressed. When evaluative uncertainty coincides with active contestation, short sellers add a narrative throughline that unifies the different lines of attack into an evocative story — an attempt to make a necessarily broad case noticeable and transmissible amid competing accounts, without giving up the coordinating function of narrative

breadth. That short sellers reserve these designs for particular environments suggests an underlying trade-off. Narrative integration makes a report more notable and easier to grasp, but it violates the conventions of the genre and can read as unprofessional. Rhetorical breadth helps audiences anticipate how differently situated others will evaluate the target, but it is cognitively demanding and exposes the report's weakest claims to attack. Focus and genre conformity therefore remain the default: short sellers appear to abandon them only when unsettled standards make second-order inference difficult, and to accept the added costs of integration only when contestation threatens the diffusion of an already unwieldy report.

Theoretical Implications

We make three contributions. First, we show that the rhetorical structure of strategic communication can be a tool not just for persuasion but also for coordination. Research on valuation has identified several ways to manage audiences with divergent standards: organizations can address them separately, rely on rankings and ratings as superordinate standards, or reconcile different standards in interaction. Each of these solutions depends on conditions that short sellers cannot rely on. Instead, they have to coordinate through the act of strategic communication itself. We show that choices that look like matters of style can serve a strategic rationale in this context. Choices like how many lines of attack a report develops and whether a narrative connects them can be used to influence what readers expect other readers to do. Short seller reports are not the only texts that are meant to solve coordination problems where individuals depend on others to act in concert with them. Proxy contests, whistleblower complaints, open letters from activist in-

vestors, and analyst initiation reports do too. Future research should explore whether the structure of these texts varies with the same conditions, or whether their coordination problems differ in theoretically relevant ways.

The findings also speak to the sociology of market efficiency. Zuckerman (2012b) argues that mispricing persists because structural barriers can block information and opinions from shaping market prices. The analysis of the strategic logic behind activist short reports clarifies one such barrier. To succeed, short sellers do not just need to synchronize dispersed actors. They need to enable them to anticipate one another's evaluative judgments. But the rhetorical architecture for this purpose work against the requirements of quick diffusion and credibility. Whether these designs succeed is a question our data do not answer. But the difficulty of the problem is itself informative: it identifies a structural reason why bubbles can persist even in the presence of informed and motivated actors, and it may help explain why activist short selling has declined as markets have become more "democratized." Retail audiences have made evaluative standards harder to anticipate and events like the GameStop rally show that short-seller's intended cascade can be turned around. Even retail investors can now organize squeezes against the positions a report discloses (Stice-Lawrence et al. 2024).

Second, we show that the strategic value of rhetorical features depends on the inference problems audiences are trying to solve. This concretizes the affordance perspective on strategic narratives, which holds that the value of narrative depends on the context in which it is deployed (Turner 2025). Specifically, we identify scope conditions for two established views on the power of narrative. The idea that stories persuade by binding many elements into one coherent whole (Rindova and Martins 2022) its settings where attention is scarce and audiences judge the story on its own terms. When audiences instead want to know whether the story can trigger actions by

audiences with other standards, a modular text with clearly separable lines of argument may serve better. Similarly, research has long held that strategic ambiguity makes a text appealing to different audiences, because each reads its own preferences into it (Eisenberg 1984; Reinecke and Donaghey 2025). Yet, ambiguity is unlikely to help when audiences are trying to infer what others think about the text because the ambiguity is itself visible. More generally, effects that have been estimated as main effects, such as the finding that elaboration raises valuations (Martens, Jennings, and Jennings 2007), may differ across audience structures. Attending to the precise inference problem that audiences of strategic communication confront, will help to improve our theories about when and why specific features of strategic communication are effective.

Third, the interviews suggest a different model of coordination in financial markets. Since Keynes, the market's inference problem has been described as third-order: each investor anticipates the conventional belief of an anonymous average. Our interviewees instead describe a cascade in which specific audiences anticipate the reactions of the specific audiences that follow — algorithms, retail investors, the institutions that read last. Where market reactions are sequenced, the third-order problem breaks into a chain of second-order problems. This makes the coordination problem of arbitrage harder than in models where investors agree on value and need only to synchronize their timing (Abreu and Brunnermeier 2003). Short sellers do not just need to synchronize actions based on shared judgments, they also need to make audience's likely judgments visible to each other.

This raises interesting questions for future research. As we saw, the first actors in short-sellers' cascades are not human but machines. Anticipating the actions of an algorithm is a different task from anticipating a human judgment. It is closer to reverse-engineering a rule than to

reading a mind. Yet, it is not clear whether the simultaneous presence of algorithms and humans makes the task of inferring likely behaviors along the cascade easier or more difficult. After all, it now becomes necessary to anticipate something about the interactions between these two entities. As automated actors take on a larger role in resolving threshold problems, it becomes important to understand whether their interactions make it easier or harder for information to be absorbed into prices. Understanding the complex interactions between interdependent and heterogenous audiences in complex valuation processes thus continues to be an important topic for research on strategic communication, valuation, and financial markets.

REFERENCES

Abreu, Dilip, and Markus K. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71(1):173–204.

Adut, Ari. 2005. "A Theory of Scandal: Victorians, Homosexuality, and the Fall of Oscar Wilde." *American Journal of Sociology* 111(1):213–48. doi:10.1086/428816.

Barber, Brad M., and Terrance Odean. 2007. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21(2):785–818. doi:10.1093/rfs/hhm079.

Beckert, Jens. 2009. "The Social Order of Markets." *Theory and Society* 38(3):245–69. doi:10.1007/s11186-008-9082-0.

Bennett, W. Lance, and Martha S. Feldman. 1981. *Reconstructing Reality in the Courtroom: Justice and Judgment in American Culture*. New Brunswick, NJ: Rutgers University Press.

Betancourt, Nathan, Inga J. Hoever, and Filippo Carlo Wezel. 2025. "Atypicality and Accountability: Evidence from Five Experiments." *Organization Science* 36(2):838–61. doi:10.1287/orsc.2022.16937.

Beunza, Daniel, and David Stark. 2012. "From Dissonance to Resonance: Cognitive Interdependence in Quantitative Finance." *Economy and Society* 41(3):383–417. doi:10.1080/03085147.2011.638155.

Boehmer, Ekkehart, and Eric K. Kelley. 2009. "Institutional Investors and the Informational Efficiency of Prices." *The Review of Financial Studies* 22(9):3563–94. doi:10.1093/rfs/hhp028.

Botelho, Tristan L. 2018. "Here's an Opportunity: Knowledge Sharing Among Competitors as a Response to Buy-in Uncertainty." *Organization Science* 29(6):1033–55. doi:10.1287/orsc.2018.1214.

Botelho, Tristan L. 2024. "From Audience to Evaluator: When Visibility into Prior Evaluations Leads to Convergence or Divergence in Subsequent Evaluations Among Professionals." *Organization Science* 35(5):1682–1703. doi:10.1287/orsc.2017.11285.

Boulongne, Romain, and Rodolphe Durand. 2021. "Evaluating Ambiguous Offerings." *Organization Science* 32(2):257–72. doi:10.1287/orsc.2020.1402.

Bourdieu, Pierre. 1993. *The Field of Cultural Production: Essays on Art and Literature*. Columbia University Press.

Bruner, Jerome. 1986. *Actual Minds, Possible Worlds*. Camrbidge, MA: Harvard University Press.

Bruner, Jerome S. 1991. "The Narrative Construction of Reality." *Critical Inquiry* 18(1):1–21.

Carroll, Glenn R., and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry." *American Journal of Sociology* 106(3):715–62.

Carruthers, B. G. 2013. "From Uncertainty toward Risk: The Case of Credit Ratings." *Socio-Economic Review* 11(3):525–51. doi:10.1093/ser/mws027.

Cattani, Gino, Simone Ferriani, and Paul D. Allison. 2014. "Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A Core-Periphery Perspective." *American Sociological Review* 79(2):258–81. doi:10.1177/0003122414520960.

Charles, Constantin, and Chad Kendall. 2024. "Causal Narratives." *National Bureau of Economic Research* No. w30346.

Charmaz, Kathy. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Pine Forge Press.

Chwe, Michael Suk-Young. 2000. "Communication and Coordination in Social Networks." *The Review of Economic Studies* 67(1):1–16.

Chwe, Michael Suk-Young. 2003. *Rational Ritual: Culture, Coordination, and Common Knowledge*. New Jersey: Princeton University Press.

COOPER, MICHAEL J., HUSEYIN GULEN, and MICHAEL J. SCHILL. 2008. "Asset Growth and the Cross-Section of Stock Returns." *The Journal of Finance* 63(4):1609–51. doi:10.1111/j.1540-6261.2008.01370.x.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage." *American Sociological Review* 82(2):297–327. doi:10.1177/0003122417691503.

Cutolo, Donato, and Simone Ferriani. 2024. "Atypicality: Toward an Integrative Framework in Organizational and Market Settings." *Academy of Management Annals* 18(1):157–209. doi:10.5465/annals.2022.0005.

Durand, Rodolphe, and Mukti Khaire. 2016. "Where Do Market Categories Come From and How? Distinguishing Category Creation From Category Emergence." *Journal of Management* 43(1):87–110. doi:10.1177/0149206316669812.

Eisenberg, Eric M. 1984. "Ambiguity as Strategy in Organizational Communication." *Communication Monographs* 51.

Elsbach, K. D., and R. M. Kramer. 2003. "ASSESSING CREATIVITY IN HOLLYWOOD PITCH MEETINGS: EVIDENCE FOR A DUAL-PROCESS MODEL OF CREATIVITY JUDGMENTS." *Academy of Management Journal* 46(3):283–301. doi:10.5465/30040623.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40. doi:10.1086/517897.

Gioia, Dennis A., Kevin G. Corley, and Aimee L. Hamilton. 2013. "Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology." *Organizational Research Methods* 16(1):15–31. doi:10.1177/1094428112452151.

Gouvard, Paul, and Rodolphe Durand. 2023. "To Be or Not to Be (Typical): Evaluation-Mode Heterogeneity and Its Consequences for Organizations." *Academy of Management Review* 48(4):659–80. doi:10.5465/amr.2020.0314.

Grodal, Stine, Michel Anteby, and Audrey L. Holm. 2021. "Achieving Rigor in Qualitative Analysis: The Role of Active Categorization in Theory Building." *Academy of Management Review* 46(3):591–612. doi:10.5465/amr.2018.0482.

Heath, Chip, Chris Bell, and Emily Sternberg. 2001. "Emotional Selection in Memes: The Case of Urban Legends." *Journal of Personality and Social Psychology* 81(6):1028–41. doi:10.1037/0022-3514.81.6.1028.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang. 2004. "Do Investors Overvalue Firms with Bloated Balance Sheets?" *Journal of Accounting and Economics* 38:297–331. doi:10.1016/j.jacceco.2004.10.002.

Hsu, Greta. 2006. "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." *Administrative Science Quarterly* 51(3):420–50. doi:10.2189/asqu.51.3.420.

Hsu, Greta, and Michael T. Hannan. 2005. "Identities, Genres, and Organizational Forms." *Organization Science* 16(5):474–90. doi:10.1287/orsc.1050.0151.

Huberman, Gur, and Tomer Regev. 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar." *The Journal of Finance* 56(1):387–96. doi:10.1111/0022-1082.00330.

Isenring, Kata, Rodolphe Durand, and Tomi Laamanen. 2025. "An Audience Heterogeneity View of Markets: Contributions, Tensions, and Agenda for Future Research." *Journal of Management* 51(6):2488–2519. doi:10.1177/01492063241312658.

Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." *American Sociological Review* 84(4):726–53. doi:10.1177/0003122419859519.

Karpik, Lucien. 2010. *Valuing the Unique: The Economics of Singularities*. New Jersey: Princeton University Press.

Keynes, John Maynard. 1936. *The General Theory of Employment Interest and Money*. New York: St. Martin's Press.

Lamont, Michèle. 2012. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38(1):201–21. doi:10.1146/annurev-soc-070308-120022.

Martens, Martin L., Jennifer E. Jennings, and P. Devereaux Jennings. 2007. "Do the Stories They Tell Get Them the Money They Need? The Role of Entrepreneurial Narratives in Resource Acquisition." *Academy of Management Journal* 50(5):1107–32.

Meléndez, Eduardo, Matthew S. Wood, and Chad Navis. 2025. "A Review of Market Categorization Research: An Evolutionary Framework Across Perspectives and Stages." *Journal of Management* 52(6):2239–73. doi:10.1177/01492063251351911.

Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3–42. doi:10.1177/0049124117729703.

Paolella, Lionel, and Rodolphe Durand. 2016. "Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market." *Academy of Management Journal* 59(1):330–51. doi:10.5465/amj.2013.0651.

Paugam, Luc, Hervé Stolowy, and Yves Gendron. 2021. "Deploying Narrative Economics to Understand Financial Market Dynamics: An Analysis of Activist Short Sellers' Rhetoric." *Contemporary Accounting Research* 38(3):1809–48. doi:10.1111/1911-3846.12660.

Phillips, Damon J., Catherine J. Turco, and Ezra W. Zuckerman. 2013. "Betrayal as Market Barrier: Identity-Based Limits to Diversification among High-Status Corporate Law Firms." *American Journal of Sociology* 118(4):1023–54. doi:10.1086/668412.

Pontikes, Elizabeth G. 2012. "Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations." *Administrative Science Quarterly* 57(1):81–118. doi:10.1177/0001839212446689.

Reilly, Patrick. 2026. "Playing to the Audiences: How Creative Producers Manage and Reconcile Different and Copresent Valuation Standards." *Organization Science* orsc.2024.18684. doi:10.1287/orsc.2024.18684.

Reinecke, Juliane, and Jimmy Donaghey. 2025. "From Constructive Ambiguity to Escalating Commitment: The Evolution of the Bangladesh Accord as a Transnational Institution for Collective Action." *Administrative Science Quarterly* 70(3):733–71. doi:10.1177/00018392251331027.

Ridgeway, Cecilia L., and Shelley J. Cornell. 2006. "Consensus and the Creation of Status Beliefs." *Social Forces* 85(1):431–53.

Rindova, Violina P., and Luis L. Martins. 2022. "Futurescapes: Imagination and Temporal Reorganization in the Design of Strategic Narratives." *Strategic Organization* 20(1):200–224. doi:10.1177/1476127021989787.

Rona-Tas, Akos, and Stefanie Hiss. 2010. "The Role of Ratings in the Subprime Mortgage Crisis: The Art of Corporate and the Science of Consumer Credit Rating." Pp. 115–55 in. Emerald Group Publishing Limited.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard university press.

Shiller, Robert J. 2019. *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Stice-Lawrence, Lorien, Yu Ting Forester Wong, and Wuyang Zhao. 2024. "Short Squeezes After Short-Selling Attacks."

Stolowy, Hervé, Luc Paugam, and Yves Gendron. 2022. "Competing for Narrative Authority in Capital Markets: Activist Short Sellers vs. Financial Analysts." *Accounting, Organizations and Society* 100:101334. doi:10.1016/j.aos.2022.101334.

Stroube, Bryan K., Keyvan Vakili, and Michaël Bikard. 2025. "The Misfit Bias." *Organization Science* 36(5):1676–89. doi:10.1287/orsc.2023.17462.

Tavory, Iddo, and Stefan Timmermans. 2014. *Abductive Analysis*. University of Chicago Press.

Than, Nga, Leanne Fan, Tina Law, Laura K. Nelson, and Leslie McCall. 2025. "Updating 'The Future of Coding': Qualitative Coding with Generative Large Language Models." *Sociological Methods & Research* 54(3):849–88. doi:10.1177/00491241251339188.

Timmermans, Stefan, and Iddo Tavory. 2022. *Data Analysis in Qualitative Research: Theorizing with Abductive Analysis*. University of Chicago Press.

Turco, Catherine, and Ezra Zuckerman. 2014. "So You Think You Can Dance? Lessons from the US Private Equity Bubble." *Sociological Science* 1:81–101. doi:10.15195/v1.a7.

Turner, Brad R. 2025. "Narrative Affordances: What Stories Can and Cannot Do." *Academy of Management Annals* 19(2):522–64. doi:10.5465/annals.2023.0101.

Zhang, X. Frank. 2006. "Information Uncertainty and Stock Returns." *The Journal of Finance* 61(1):105–37. doi:10.1111/j.1540-6261.2006.00831.x.

Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104(5):1398–1438. doi:10.1086/210178.

Zuckerman, Ezra W. 2004. "Structural Incoherence and Stock Market Activity." *American Sociological Review* 69(3):405–32. doi:10.1177/000312240406900305.

Zuckerman, Ezra W. 2012a. "Construction, Concentration, and (Dis)Continuities in Social Valuations." *Annual Review of Sociology* 38(1):223–45. doi:10.1146/annurev-soc-070210-075241.

Zuckerman, Ezra W. 2012b. "Market Efficiency: A Sociological Perspective." in *Oxford Handbook of the Sociology of Finance*, edited by K. Knorr-Cetina and A. Preda. New York: University of Oxford Press.

Zuckerman, Ezra W. 2017. "The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool." Pp. 31–68 in *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads*. Emerald Publishing Limited.

TABLES AND FIGURES

Table 1. Descriptive Statistics of the Activist Short Seller Field.

Short-seller profile	Firms	Campaigns	Campaigns per seller, M (SD)	Share of corpus	Median sectors targeted	Median top-sector share
Generalists (≥2 campaigns, <50% in one sector)	36	599	16.64 (17.42)	72.2%	6	33%
Specialists (≥2 campaigns, ≥50% in one sector)	40	183	4.58 (5.08)	22.0%	2	62%
One-shot authors (1 campaign)	48	48	1.00 (0.00)	5.8%	1	100%
All short sellers	124	830	6.69 (11.73)	100.0%	2	67%

Note: Coded corpus N = 830 campaigns by 124 distinct short sellers, 2008–2024, across 22 2-digit NAICS sectors.

Table 2. Data Sources, Variables, Coverage.

Variable	Source	Definition	N	Mean (SD)	Coverage
Outcomes (abnormal returns)					
CAR, [−1,+1]	CRSP	Market-model cumulative abnormal return in the [−1,+1] window.	746	-0.074 (0.139)	90%
CAR, [0,+1] (announcement)	CRSP	Market-model CAR in the [0,+1] window.	746	-0.067 (0.119)	90%
CAR, [0,+21] (drift)	CRSP	Market-model CAR over the post-report month — the drift window.	746	-0.113 (0.290)	90%
CAR, [0,+252] (one year)	CRSP	Market-model CAR over the year after the report.	740	-0.608 (1.41)	89%
Target characteristics					
Log market capitalization	CRSP	Log market value of equity on the last trading day before the campaign.	789	20.86 (1.63)	95%
Industry (2-digit NAICS)	Compustat / CRSP	The target's 2-digit NAICS sector, used for industry fixed effects.	829	22 sectors	100%
Book leverage	Compustat	Total debt over total assets, from the latest quarter.	762	0.060 (0.207)	92%
Asset growth	Compustat	Year-over-year growth in quarterly total assets (Cooper, Gulen, and Schill 2008), from the latest pre-announcement quarter.	731	2.17 (8.04)	88%
Net operating assets	Compustat	Net operating assets scaled by total assets (Hirshleifer et al. 2004), from the latest pre-announcement quarter.	759	0.618 (1.69)	91%
Context measures					
Evaluative uncertainty (EUS)	CRSP / Compustat / IBES / 13F / RavenPack	How hard the firm is to value: an index of thin analyst coverage, concentrated ownership, a short public history, and little media attention.	809	0.503 (0.141)	97%
Pre-campaign volatility (60-day)	CRSP	Realized volatility of daily log returns over the 60 trading days before the campaign.	780	0.801 (0.578)	94%
Analyst forecast dispersion	IBES	Spread of analyst target-price forecasts (coefficient of variation) before the campaign.	508	0.197 (0.156)	61%
Report features					
Report length (segments)	Report (Stage-01 extraction)	Number of semantic segments the report was split into.	830	25.91 (29.14)	100%
Report length (words)	Report (Stage-01 extraction)	Word count of the report body.	830	6,485 (7,918)	100%
Report length (pages)	Report (Stage-01 extraction)	Page count of the source report.	830	29.06 (33.97)	100%
Qualitative layer					
Audience-reaction sources	RavenPack/DJ news; Reddit/Stocktwits; CIQ transcripts; Refinitiv sell-side; X / Seeking Alpha	Reaction capture per case across press, retail, institutional, and social channels: ≈17,600 retail posts, ≈3,700 newswire stories, ≈1,000 call transcripts and sell-side hits, plus per-case web supplements.	44		≈22,500 items
Practitioner interviews	Author-conducted	Semi-structured interviews with activist short sellers and financial journalists who cover the field.	21		≈60 min avg.

Table 3. Higer-order constructs: Definitions and Illustrative Examples.

Category	Values	Definition	Example
Tone	Professional	Sober, plain reporting, with at most an isolated rhetorical flourish.	Accounts receivable grew 3x faster than revenue over eight quarters. The company's current ratio was a deeply worrying 0.35, down from 0.59 at the end of 2018.
	Colorful	Some vivid or dramatic language, but not a sustained performance.	Has anyone ever heard of a company voluntarily giving up money in hand for a questionable receivable? This lawsuit lays out a rationale that describes how UOP is not only trying to deceive their clients, but more importantly, the Department of Education.
	Theatrical	A clear performance throughout — dramatic reveals, mockery, insider winks, or hyperbole.	THIS IS THE LAWSUIT APOLLO GROUP MANAGEMENT DOES NOT WANT THE GOVERNMENT OR INVESTORS TO SEE.
Narrative structure	Narrative	Organizes events into a story — actors, happenings, and language linking them over time.	June 9, 2016 - OSIR President and CEO Dwayne Montgomery resigns suddenly for "medical reasons". His replacement, David Dresner, has no healthcare experience. During the period September 2015 to June 2016, the company churned through three CEOs and two CFOs.
	Pure argument	States claims, evidence, or description without arranging them into a causal sequence.	We have no strong bearish or bullish view on the future of blockchain. We genuinely hope the technology is implemented broadly and that currency and information can be effectively decentralized through its use. Regardless of ones views on blockchain technology however, we think Riot is a name that investors should avoid. We urge cautious investing to all.
Narrative trope	Boom and Bust	Blames the market: a whole sector caught in cyclical excess or mania that is bound to unwind.	Blucora, formerly known as Infospace, is one of the truly great historical case studies of the fraud and excesses committed during the dot com boom of the late 1990s. Former Merrill Lynch analyst Henry Blodget privately referred to Infospace as junk, even as he publicly remained bullish. Blodget was subsequently indicted for securities fraud, and agreed to a permanent ban from the securities industry.
	Structural Risk	Blames the industry: outside forces like obsolescence or secular decline doom the business.	entrotech and PPG have effectively integrated protection technology directly into OEM paint, hence disintermediating XPEL's aftermarket solution. The technology is already being rolled out among the Big 3 automakers with the anticipation that within the next 5 years, the entire aftermarket industry will be disrupted.
	Shaky Foundations	The company's own business model structurally leads to failure, whoever runs it.	Lee Enterprises has embarked on an unsustainable business model of destroying its product while raising prices... leverage already caused the company to file for bankruptcy once before, in 2012. While this strategy has been successful over the past few years, it is unlikely to be successful going forward. LEE's strategy is doomed to fail... financial performance will deteriorate once consumers catch on and the company runs out of cost cutting measures.
	Smart money vs dumb money	Informed insiders are quietly getting out while naïve outsiders keep buying.	Since the public release of the Zilretta diabetes data on November 1, 2016, insiders have been sellers of stock. Former Chief Medical Officer Bodick Neil has sold more than $2.4M in stock, starting almost immediately after the data was released and progressing at an unrelenting pace.

Category	Values	Definition	Example
	Stock Hype	Someone is actively promoting the stock, inflating expectations beyond the fundamentals.	In the case of Terra Tech, it crawled out of the great stock promotion primordial ooze. It all started with attorney Thomas Puzzo, who has set up shells for promotions over and over. All of these promotions resulted in shareholder losses of at least 90%. Puzzo followed a similar pattern to set up the shell that eventually became TRTC.
	Bumbling Fools	Blames management incompetence: the business could work under better leadership.	Served as CEO of Davnet, an Australian-listed company which collapsed with the dotcom bubble due to massive liquidity issues. It turned out that Hal Turner had excellent money-dissipating skills. When Turner had been in charge in 2000-01 there was much going on in the company. Davnet's asset position fell by $100 million. At the same time, its cash reserves shrank by $72 million, leaving about $11 million in the once-bursting company coffers.
	Hidden Skeletons	The company is deliberately concealing or burying material information.	In each case, although UOP had already received the students' loan proceeds, they refunded 100% of it to the government, and opted instead for the far less lucrative path of trying to collect directly from the student. By removing these early-withdrawing students from their loan rolls, this lawsuit suggests understatement of Cohort Default Rates, plus other impacts on enrollments and revenue ratios.
	Hustler	Casts insiders as immoral actors — fraudsters or manipulators enriching themselves.	Blink's business isn't designed to ever generate cash, only to continually stuff the pockets of insiders. Blink mirrors Farkas's involvement in multiple previous criminal schemes. The stock collapsed, and Skyway's principals were sued by the SEC for the pump-and-dump scheme.
	Impossible Challenge	The company faces obstacles it is unlikely to overcome given its resources.	Vicor cannot deliver product. Vicor is building a new facility to do the electroplating but given the challenges without sourcing of high current parts and given the typical 12 months it takes to ramp up production, they face significant challenges in meeting customer demands.

Note: Full definitions and passages are in Appendix E

Table 4. Reliability of the Coded Constructs Underlying the Content and Form Typologies.

Construct	Coding unit	How the analysis uses it	Metric	*Agreement on that quantity*		*Campaigns*	
				vs. GT	**Among models**	**GT**	**Corpus**
Trope presence, binary (9 tropes, pooled)	Segment × trope → campaign	Content-cluster input (9 binary presence flags), for rhetorical breadth indicator	Cohen's κ	.569	.693	37	822
Narrative share (narrative_pct)	Segment → campaign	Corpus-level indicator of narrative integration; association with active contestation within broad reports; configuration map (Figure 5)	Spearman ρ	.509	.646	38	829
Tone (tone_avg)	Segment → campaign	Continuous descriptive form measure; house-style decomposition (Table 5)	Pearson r	.801	.792	38	829

Note: Narrative share is used as a continuous corpus-level indicator of narrative integration and is scored as a rank by Spearman's ρ and the mean pairwise ρ. All agreement statistics are computed at the campaign level against the adjudicated ground-truth pool of 37–38 campaigns; the full audit, including inter-model agreement and level bias, appears in Appendix E (Table E3). Campaign trope breadth, the count underlying the broad/focused split, tracks the adjudicated ground truth at Spearman ρ = .88 under the majority rule and ρ = .90 under unanimity.

Table 5. House Style and Its Limits: Between-Seller Variance Components of Report Features.

Report feature	Type	Between-seller share (η^2)
Tone	Rhetorical Form	.505
Trope Breadth vs Focus	Content Repertoire	.302
Narrative share	Rhetorical Form	.231

Note: N = 685 campaigns across 40 short sellers with ≥ 5 campaigns each (the between/within split is undefined for one-off sellers, which would have zero within-seller variance and inflate the signature). Each entry displays the share of a feature's variation that lies between sellers rather than within a seller's own campaigns (one-way ANOVA grouped by seller).

Table 6. Rhetorical Architecture and the Evaluative Environment: OLS Regression Estimates

Variable	(1) Unadjusted	(2) Core controls	(3) Full controls	(4) + Seller fixed effects
Panel A. Dependent variable: Broad report (linear probability)				
Evaluative uncertainty (standardized)	.078**	.068**	.067*	.073**
	(.027)	(.026)	(.028)	(.023)
Book leverage		−.007	.027	.105
		(.070)	(.100)	(.104)
Asset growth			−.006***	−.002
			(.001)	(.002)
Net operating assets			.021	.077
			(.049)	(.050)
Announcement-year fixed effects	No	Yes	Yes	Yes
Two-digit NAICS fixed effects	No	Yes	Yes	Yes
Short-seller fixed effects	No	No	No	Yes
Observations	809	750	680	680
Short-seller clusters	121	116	114	114
R^2	.029	.092	.089	.403
Panel B. Dependent variable: Narrative share (0–1)				
Pre-campaign volatility (standardized)	−.010	−.011	−.004	−.020
	(.014)	(.015)	(.016)	(.015)
Broad report	.040	.034	.030	.022
	(.024)	(.025)	(.026)	(.021)
Volatility × broad report	.043*	.045*	.039*	.042*
	(.017)	(.018)	(.020)	(.018)
Book leverage		.013	.006	.015
		(.025)	(.030)	(.024)
Asset growth			.001	.001*
			(.000)	(.000)
Net operating assets			.003	−.007
			(.027)	(.028)
Announcement-year fixed effects	No	Yes	Yes	Yes
Two-digit NAICS fixed effects	No	Yes	Yes	Yes
Short-seller fixed effects	No	No	No	Yes
Observations	780	720	654	654
Short-seller clusters	118	113	111	111
R^2	.037	.071	.063	.427

Note: Entries are OLS coefficients with short-seller-clustered standard errors in parentheses; all tests are two-tailed. The models are estimated in the direction the theory specifies — from target environment to report design. Panel A is a linear-probability model of the broad-report indicator on evaluative uncertainty standardized on its full available sample; the unadjusted coefficient corresponds to the raw broad-minus-focused gap of 0.37 standard deviations of evaluative uncertainty (raw means .519, SD .142, for broad and .466, SD .131, for focused reports; Welch $t(472.49) = 5.08$, $p < .001$). Panel B regresses narrative share (proportion of narrative segments, 0–1) on standardized 60-day pre-campaign volatility, the broad indicator, and their interaction; the volatility coefficient is the slope among focused reports, and the interaction is the broad-minus-focused slope difference. In the unadjusted data, narrative share correlates .24 with volatility among broad reports ($p < .001$) and −.04 among focused reports ($p = .496$). The reversed models, with the environment measures as outcomes, yield the same pattern (interaction .114 unadjusted to .060 fully controlled per 10 pp of narrative share, $p < .05$ throughout), but only the architecture-as-outcome interaction shown here survives short-seller fixed effects — column (4) — so those estimates are identified within seller: the same seller writes

broader on higher-uncertainty targets and more narratively when a broad report targets a more volatile firm. Core models include announcement-year and two-digit-NAICS fixed effects plus book leverage; full models add asset growth and net operating assets; column (4) adds a fixed effect for each short seller (R^2 rises mechanically). NAICS categories with fewer than 10 corpus campaigns are pooled. Models use specification-specific complete cases. Firm size and listing age are omitted because they closely proxy evaluative uncertainty itself; report length is assessed separately as a mechanical rival. * p < .05, ** p < .01, *** p < .001. Associations, not causal estimates.

Table 7. Illustrative Cases of the Three Report Strategies.

Case	Strategy	EUS · Vol	Fit	How the campaign works as its strategy	Outcome
Low Evaluative Uncertainty — typical strategy: Focused					
Transdigm Group Citron Research · 2017 ✓	Focused	low · low	Fit	Compresses the entire case into a single analogy — the aerospace-parts maker as 'the Valeant of the Aerospace Industry' — pegged to the incoming administration's promise to cut military procurement, which would expose the firm's reliance on serial price hikes.	Success (-4%, -0.53σ)
Axos Financial Real Talk Investments · 2015 ✓	Focused	low · high	Fit	Develops one narrative of corporate obfuscation: that the transcript the bank filed as an 8-K exhibit is 'materially different than the BOFI Transcript' of its own investor call.	Success (-24%, -0.88σ)
Applied Digital Corporation The Friendly Bear · 2023 ✓	Focused	low · high	Fit	Builds a single argument in a tight legal-forensic register, focused on 'hidden skeleton' tropes: the AI-datacenter highflyer is secretly a controlled company of its patron investment bank.	Success (-69%, -1.31σ)
Axos Financial The Friendly Bear · 2015 ✓	Integrated	low · low	Misfit	Follows a New York Times investigation by arguing the branchless bank's peer-beating margins rest on risky mortgages to 'unsavory' borrowers — then layers further narratives of compliance failures, undisclosed self-dealing, and weak regulatory standing around that core claim.	No move (+4%, +0.50σ)
Burlington Stores Spruce Point Capital · 2016 ✓	Modular	low · low	Misfit	Stacks a self-described 'plethora' of accounting red flags across a fifty-nine-slide deck, each headed 'Warning:', with no single story or argument dominating the presentation.	No move (+13%, +2.29σ)
High Evaluative Uncertainty · Limited Contestation — typical strategy: Modular					
Endurance International Group Gotham City Research LLC · 2015 ✓	Modular	high · low	Fit	Assembles several separable arguments and narratives — revenue quality read from the roll-up's own disclosures against irreconcilable Indian regulatory filings, subsidiary fraud, reputational and regulatory exposure from its customers, and an unsustainable debt load — each resting on its own evidence rather than a single thesis.	Success (-8%, -0.65σ)
PetIQ Spruce Point Capital · 2019 ✓	Modular	high · low	Fit	Combines a genealogical exposé of the executives, grey-market drug diversion, rebate-dependent earnings, a cash-burning clinic rollout, and a plain peer valuation. Each argument or narrative offers an independent reason to short.	Success (-7%, -0.59σ)
Weis Markets Spruce Point Capital · 2018	Modular	high · low	Fit	Stacks separable forensic flags — its own price checks and credit-card-panel data, auditor turnover, and an accounting choice that counts 'funds in transit under 7 days' as cash — to argue a grocer is masking financial strain.	Success (-8%, -1.23σ)
Runway Growth Finance Corp NINGI Research · 2023 ✓	Integrated	high · low	Misfit	Several separable lines of argument (understated loan values, dividends paid out of capital, and marked-up asset values at a business-development lender), but heavily relies on narratives about distress at different portfolio companies to tie the report together.	No move (+5%, +0.66σ)

Case	Strategy	EUS · Vol	Fit	How the campaign works as its strategy	Outcome
Trupanion Long Short Value · 2017	Integrated	high · low	Misfit	Narrates a focused overvaluation story around the pet insurer's 'razor thin insurance margins' and 'no real moat outside of brand name.'	No move (-2%, -0.24σ)
High Evaluative Uncertainty · Active Contestation — typical strategy: Integrated					
Clearside Biomedical Cliffside Research · 2016 ✓	Integrated	high · high	Fit	Concedes the bull case up front — the biotech is 'a fine company' likely to win approval — then pulls a re-derived patient count, a cautionary drug history, and insider red flags into a single chronological arc that ends with a clear prediction: insiders already own the good news and will exit when shares come off lock-up and early sales disappoint.	Success (-90%, -2.57σ)
Nikola Hindenburg Research · 2020 ✓	Integrated	high · high	Fit	Threads a wide set of revelations through the founder's biography — recorded calls, texts, and a video of a truck rolling downhill — into one overarching character narrative of lifelong deception that ends by asking whether he is a visionary or 'a con artist.'	Success (-62%, -1.34σ)
Intelligent Systems Marcus Aurelius Value · 2019 ✓	Integrated	high · high	Fit	Fuses a governance exposé of a named director — the disgraced former chief executive of MiMedx installed as the audit committee's financial expert — with a re-model of the fintech as 'largely a niche Indian outsourcing business,' binding the two lines into a single story of concealment.	Success (-56%, -1.96σ)
Control4 Cannell Capital · 2016 ✓	Integrated	mid · mid	Fit (boundary)	Ties a variegated set of complaints — financial problems, unfavorable product comparisons and reviews, insider share sales, board resignations, potential stock hype — together with an obfuscation narrative, concluding the company 'hides behind adjustments and acquisitions to obscure' its lack of profits.	No move (+19%, +0.91σ)
Purecycle Technologies Holdings Hindenburg Research · 2021	Modular	high · high	Misfit	Attacks the plastics-recycling SPAC on several argumentative fronts — 'the latest zero-revenue, ESG-themed SPAC,' executives with a record of 'destroying public shareholder capital,' and a licensed process a polymer expert calls 'a bomb' — without binding them under a single overarching narrative.	No move (-6%, -0.23σ)

Note: Illustrative cases drawn from a larger set of close read comparisons; the complete set of coded cases appears in Appendix G. Cases are grouped by target environment (bold rows). Within each, 'Fit' marks cases whose strategy matches the environment mapping established in the quantitative analysis; 'Misfit' marks cases whose strategy departs from it. The cases illustrate how each strategy operates; they are not evidence that fit produces returns — the corroborating fit–return estimates appear in Table D3. Outcome reports the abnormal decline (CAR) over the post-report month. The checkmarks in the case column mark which cases are discussed in the paper.

Figure 1. Example of two short reports.

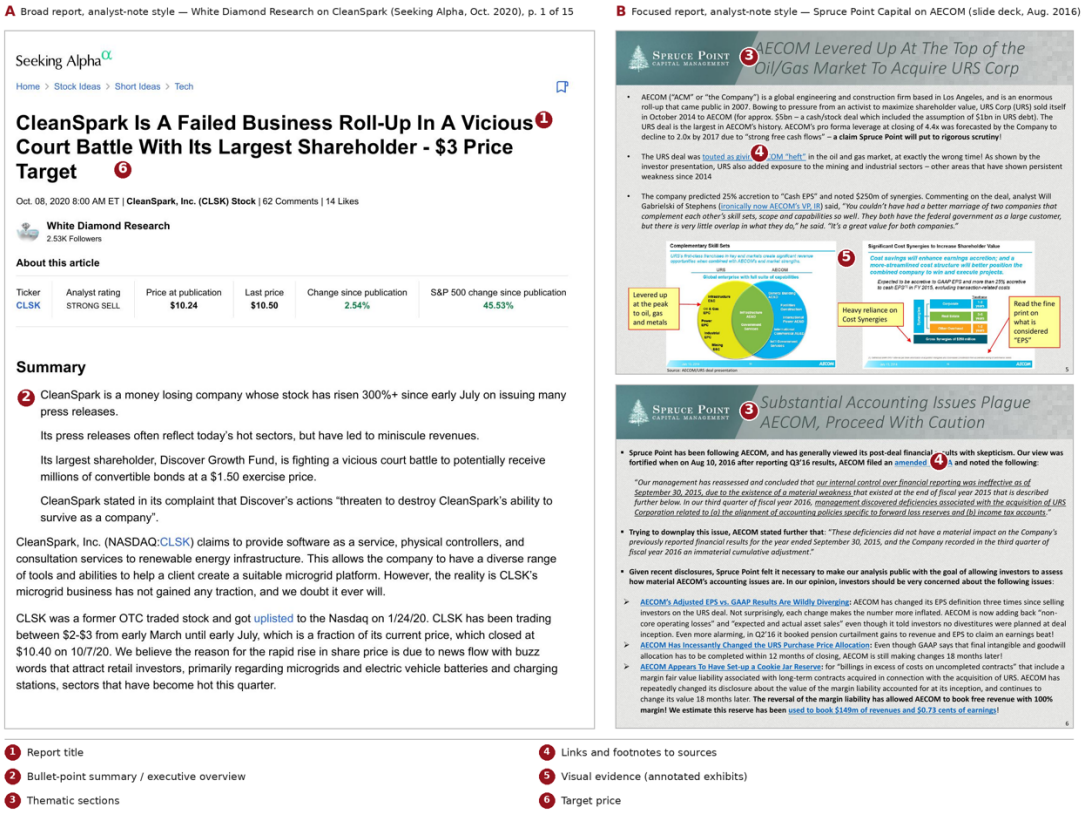
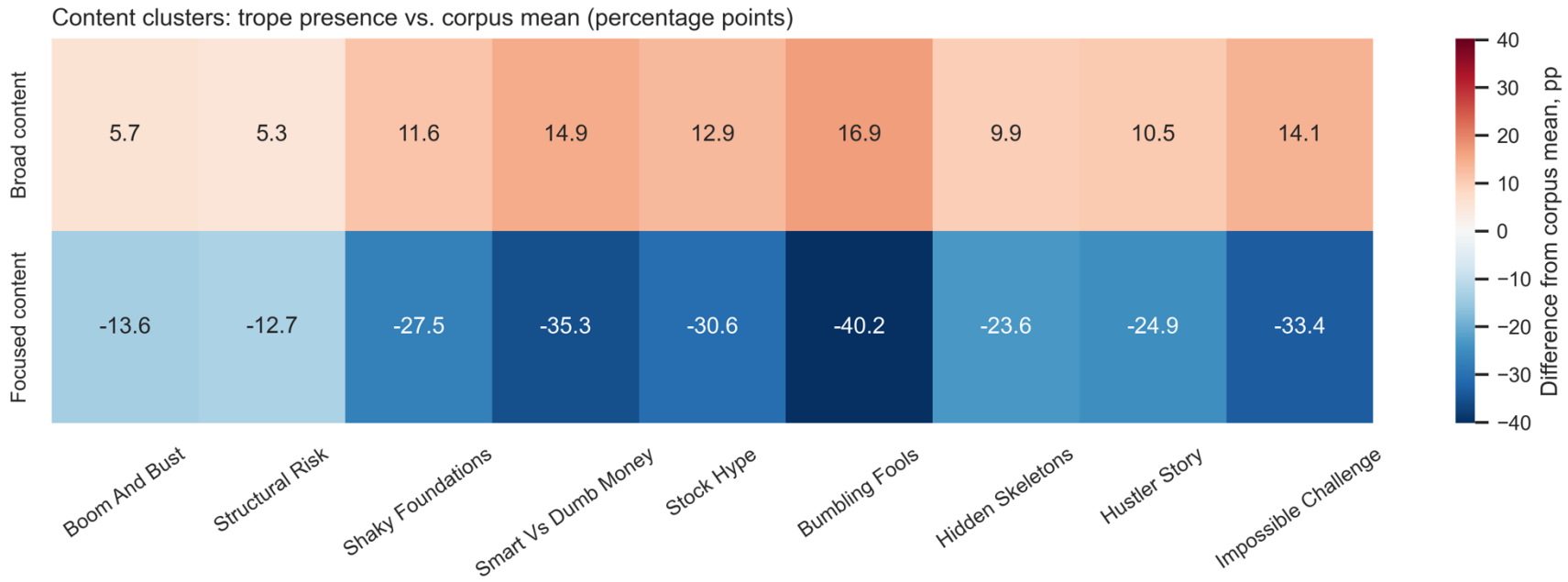


Figure 2. Heatmap of Content and Form Cluster Composition.



Note: Positive cells mark over-represented dimensions. The nine binary trope-presence rates by content cluster, as percentage-point differences from the corpus mean.

Figure 3 – Distribution of Reliance on Narrative and Tone Variables.

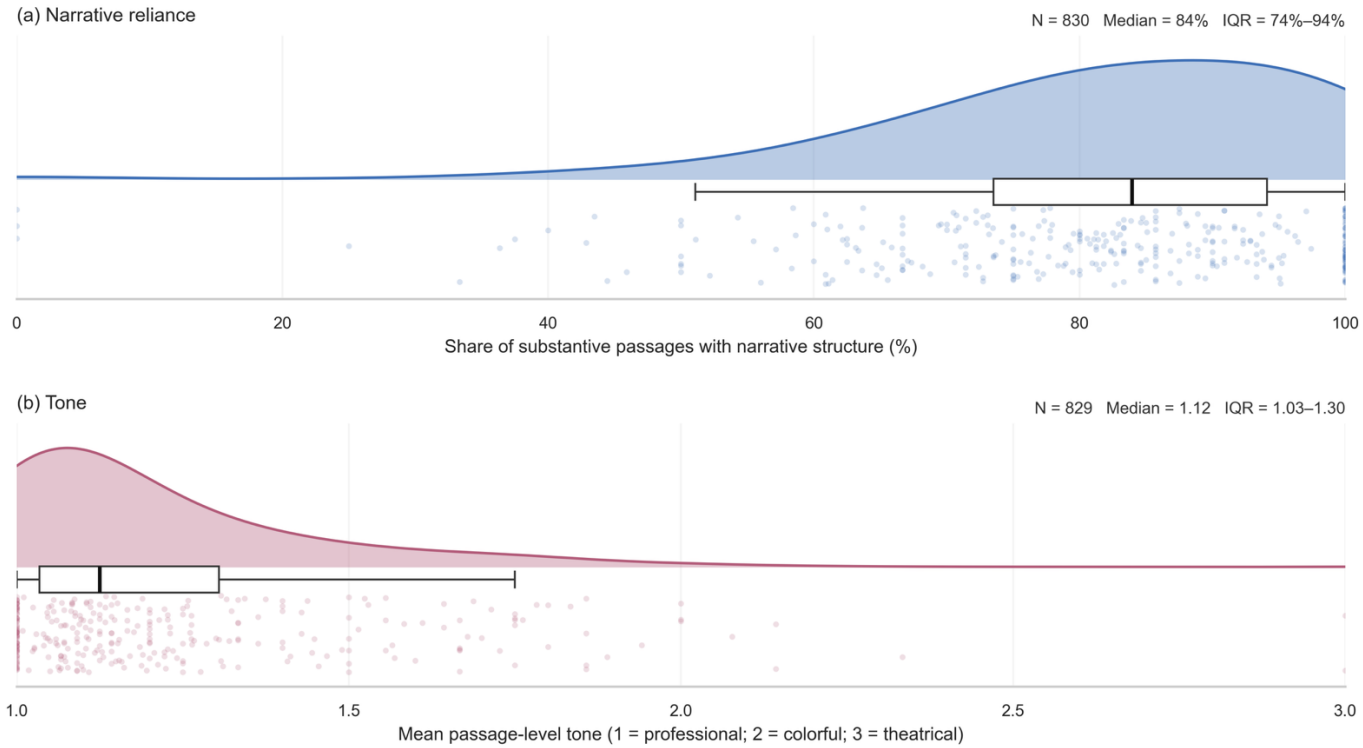
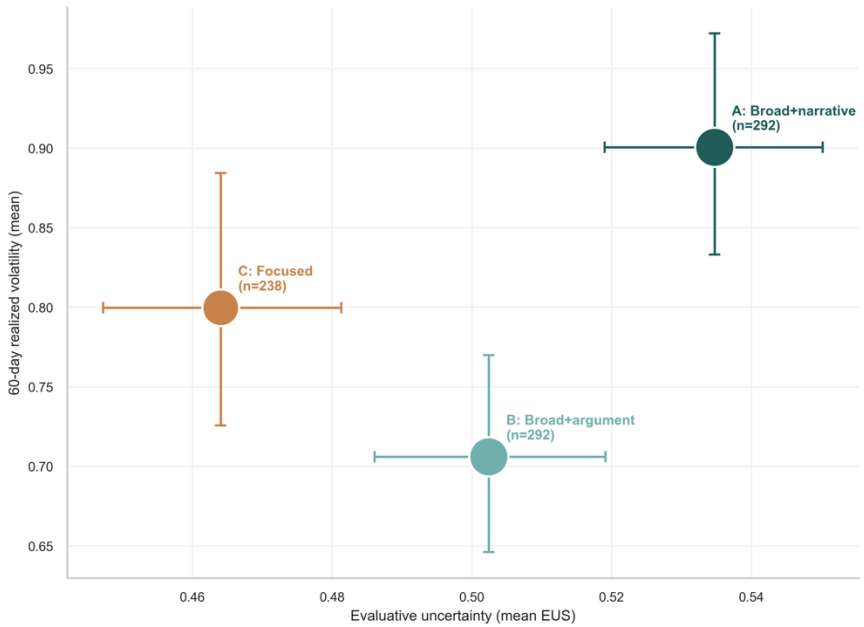


Figure 4. The Three Strategies, mapped to Volatility and Evaluative Uncertainty.



Note: Points are the strategy means of evaluative uncertainty and 60-day pre-campaign realized volatility, with bootstrap 95% confidence intervals. Broader reports target harder-to-value companies (partial R^2 = .019**, N=680) and, within broad reports, a more narrative register accompanies more contested, volatile targets (.038**, N=452), net of year, industry, leverage, and target fundamentals, with standard errors clustered by short seller.

ONLINE APPENDIX

Appendix A — Variable construction

Most variables are constructed according to standard procedures and are sufficiently described in the data section. However, our measure of evaluative uncertainty was constructed by hand and is sufficiently complex to warrant further elaboration. Furthermore, to operationalize active disagreement with volatility required some further robustness tests described below.

Evaluative Uncertainty (EUS). Leaning on Botelho (2018), we build our uncertainty variable from five basic components that measure the degree of scrutiny a company faces from key institutions in the market. Keeping with prior research (Boehmer and Kelley 2009; Zhang 2006), the underlying assumption is that more intense scrutiny and longer coverage translates into greater certainty about how the company should be evaluated. We use the variable with some caution as it conflates scrutiny by knowledgeable players with agreement about the correct evaluative standard. This is often but not always appropriate.

Firm Age was calculated as the difference between the short seller campaign and the initial year the stock was covered in the Compustat/CRSP database - typically the day of the IPO. Firm age proxies the general availability of information about the firm. The older it is, the more certain the general knowledge about strategy, leadership and performance has become.

Sell-side coverage is the sum of I/B/E/S NUMEST analyst counts over the four most recent pre-announcement coverage quarters. Statistical summaries dated on the campaign day are excluded. A matched observation beyond the prespecified staleness window is treated as zero active coverage, while firms with no valid identifier match remain missing. Greater active coverage indicates more sustained scrutiny and a more widely shared evaluative apparatus.

Media coverage is the number of distinct Dow Jones/RavenPack stories in the 30 calendar days before the campaign, excluding the campaign date, with entity relevance of at least 90; the Botelho-style input adds one to this count. A thinner pre-campaign media environment indicates less outside scrutiny and therefore greater evaluative uncertainty.

We captured institutional ownership with two separate measures, based on 13-F holdings retrieved from Thomson Reuters. On the one hand, research has found that highly concentrated ownership among institutional investors limits the circulation of information - and thus the recursive interplay that stabilizes classification schemas. It therefore tends to increase evaluative uncertainty (Boehmer and Kelley 2009; Botelho 2018). On the other hand, broad institutional ownership reduces uncertainty by increasing sophisticated scrutiny, relative to other owners. To capture both dimensions, we included institutional ownership concentration as the ratio of the sum of shares of a firm's stock that are owned by each of the five largest institutions to the total number of shares outstanding; and the proportion of shares held by all remaining institutional investors. Following Botelho, the intuition behind these two measures is that high concentration among the top five investors reduces the amount of data that circulates in the broader economy, effectively increasing evaluative uncertainty. Conversely, broad diffusion of ownership among many institutional investors means an increase in scrutiny from sophisticated actors. Because we wanted to use the measure for within-sample comparisons of rhetorical strategies, we oriented each of the five components so that higher values indicate greater uncertainty, mapped each to its percentile rank within the scoreable sample, and took the equal-weighted mean of the available percentile ranks as our final, continuous measure of evaluative uncertainty (requiring at least four of the five components, which yields scores for 809 of the 830 coded campaigns and 849 of all

873 master-file campaigns). Though this approach means that we cannot make absolute statements about evaluative uncertainty, it gives us an intuitive indication where to expect, relatively speaking, more or less evaluative uncertainty in our sample. In our qualitative analysis, we used the subscores individually and alongside the composite measure to better contextualize evaluative uncertainty in specific campaigns.

Pre-campaign volatility. The 60-day pre-campaign volatility measure is constructed from CRSP daily data obtained through WRDS. For each campaign, we compute daily log returns, $\log(1 + r)$, over a window of 60 trading days ending strictly before the campaign announcement (the last day of the window is $t-1$). The measure itself is the standard deviation of these daily log returns, annualized by the square root of 252. Campaigns with fewer than 20 valid trading days in the window are set to missing, and windows containing at least 75 percent of the target days are flagged as good quality. As robustness checks, we compute the Garman–Klass (OHLC-based) and Parkinson (high–low range) estimators over the same window, and repeat the construction at 30-, 90-, and 180-day horizons; the same script also produces average relative trading volume (daily volume scaled by its prior 60-day rolling mean), which serves as the turnover control in the analyses. Substantively, this measure serves as the second feature of the target's evaluative environment — disagreement realized in prices — as distinct from evaluative uncertainty. We validate this interpretation by showing that volatility tracks analyst forecast dispersion, a direct measure of disagreement, at $\rho \approx 0.52$ raw and ≈ 0.49 after partialling out target size, turnover, and the evaluative uncertainty score itself, while its correlation with the uncertainty score is only about 0.40. This is reported in Table A1. Dispersion itself is only weakly related to evaluative uncertainty (Spearman $\rho = .16$ raw) and unrelated once size and turnover are

partialled out (ρ = .06, p = .21): the direct disagreement measure tracks the contestation dial specifically, not uncertainty.

Table A1. Volatility as Realized Contestation: Analyst-Dispersion Validation Battery.

Relationship	Rank ρ	p	N
EUS × volatility (raw)	.40***	< .001	815
Volatility × analyst dispersion (raw)	.52***	< .001	527
Volatility × dispersion, net of size + turnover	.49***	< .001	527
Volatility × dispersion, net of size + turnover + EUS	.49***	< .001	527

Appendix B — Interviews

Table B1. Interview List.

#	Date of interview	Pseudonym	Professional experience	Years of experience
1	08/05/2024	Nick	Partner, long/short hedge fund	10
2	08/07/2024	Blake	Financial journalist	30
3	08/27/2024	Rumi	Financial journalist, short-focused analyst	30
4	09/23/2024	Roger	Financial journalist, short-focused analyst	40
5	09/24/2024	Luke	Financial journalist, short-focused analyst	50
6	09/26/2024	River	Financial journalist	40
7	02/18/2025	Charlie	Principal, short-focused hedge fund	10
8	02/28/2025	Thomas	Principal, short-focused hedge fund and short activist	15
9	02/28/2025	Don	Principal, short-focused hedge fund and short activist	20
10	03/05/2025	Winter	Principal, short-focused research firm	5
11	03/06/2025	Arthur	Principal, short-focused hedge fund and short activist	15
12	03/07/2025	Otto	Principal, short-focused hedge fund and short activist	5
13	03/09/2025	Sam	Principal, short-focused hedge fund and short activist	20
14	05/13/2025	Jackie	Principal, short-focused hedge fund	40
15	06/05/2025	Zeke	Principal, short-focused hedge fund and short activist	20
16	06/16/2025	Saul	Principal, short-focused hedge fund and short activist	30
17	04/07/2026	Riley	Principal, short-focused hedge fund	20
18	06/29/2026	Jules	Principal, short-focused hedge fund and short activist	15
19	07/07/2026	Quinn	Principal, short-focused hedge fund and short activist	20
20	07/08/2026	Bryce	Principal, short-focused hedge fund and short activist	20
21	07/08/2026	Marion	Principal, short-focused hedge fund and short activist	10

Note: Sixteen semi-structured interviews conducted between August 2024 and July 2026 with short sellers, short-focused analysts, and financial journalists. All names are pseudonyms, and professional experience is described at the level of detail that preserves anonymity.

Appendix C — Clustering procedure and robustness

This Appendix describes the clustering procedure in greater detail. We cluster campaigns at the content level on the nine binary trope-presence indicators (whether each trope appears anywhere in the campaign under the majority vote rule). Binary presence is more defensible than salience rates as a clustering input because it asks only whether a trope appears at all, not how many segments support it — an aggregation that is robust to segment-level noise (see Appendix E for details). We sweep five different clustering algorithms — k-means with fifty random initializations; Ward agglomerative hierarchical clustering with Euclidean distance; average-linkage hierarchical clustering with Hamming distance (appropriate for binary features); spectral clustering with a k-nearest-neighbors graph; and a soft-balanced k-means variant — and we sweep across k from two to six. For each candidate solution we compute six diagnostics: silhouette, Calinski-Harabasz, Davies-Bouldin, minimum cluster share, maximum cluster share, and the size ratio between the largest and smallest clusters. We then re-estimate the solution on one hundred bootstrap resamples drawn without replacement at 80% of the original sample size and compute the adjusted Rand index between the original-sample labels and the resample labels.

We then apply these four criteria jointly to identify a set of potentially defensible solutions. Specifically, a candidate solution must reach a silhouette score of at least 0.25, place at least 5% of campaigns in its smallest cluster, keep the ratio of largest to smallest cluster below twenty, and achieve a mean bootstrap-adjusted Rand index of at least 0.70 across resamples. The four criteria address conceptually distinct properties — geometric separation, substantive balance, size symmetry, and sample-perturbation stability — and a solution must clear all four. We treat the four-criterion screen as the bar for calling a solution a defensible cluster solution, rather than a useful descriptive partition.

Among these defensible solutions, we then select the ones with the best tradeoff between stability and separateness. The primary content solution is k-means at k = 2. The solution has a silhouette score of 0.26, its minimum cluster share is approximately 30%, the size ratio is approximately 2.4, and the bootstrap mean adjusted Rand index is approximately 0.92. The hierarchical-Ward adjusted Rand index against the same partition is 0.49.

The content cluster divides into a focused and a broad configuration. The broad-versus-focused contrast is stable across vote rules: the same k = 2 split reproduces under the lenient four-voter rule that incorporates the Llama audit voter. It is, however, specific to the standardized-Euclidean k-means geometry — average-linkage clustering with a Hamming distance does not recover it, instead isolating a small, sparse-repertoire tail of campaigns in an unstable partition that fails the defensibility screen. Under the strict unanimous rule a two-way broad-versus-focused split is still recoverable, but it is more balanced and falls below the silhouette separation threshold, so we treat it as a descriptive rather than a defensible partition. Trope clustering based on counts rather than binary presence does not clear the defensibility screen at any k, because the rate variables are noisier and more length-sensitive than presence indicators. Mixed-layer feature sets (tropes plus form) yield well-balanced but poorly separated partitions that fall below the silhouette threshold, so we keep the content and form layers distinct.

Appendix D — Corroborating EUS / Volatility Associations with Report Structure

The analysis shows that the rhetorical breadth of a report tracks evaluative uncertainty about the target, while the share of narrative versus argumentative passages tracks volatility. Figure 4 in the paper illustrates both associations but does not report the underlying statistics. Neither association requires elaborate machinery, and Table D1 therefore reports them in elementary form: one comparison of group means and three correlations. Broad reports target companies whose evaluative uncertainty is 0.37 standard deviations higher than the targets of focused reports (Welch $t = 5.08$, $p < .001$); expressed as a point-biserial correlation, $r = .17$, so the content split accounts for about 3 percent of the variance in evaluative uncertainty ($r^2 = .029$). Narrative share correlates with pre-campaign volatility at $r = .24$ ($p < .001$) within broad reports and at $-.03$ ($p = .62$) within focused ones; an ordinary least-squares interaction confirms that the two slopes differ ($\beta = +1.12$, $t = 4.20$, $p < .001$). The continuous nine-trope count alone barely correlates with evaluative uncertainty ($\rho = .05$, $p = .14$). The association is carried by membership in the broad versus focused repertoire family — which bundles breadth with the absence of a dominant storyline — not by the raw count, which is additionally suppressed by report length: partialling out log report word count raises the breadth–uncertainty correlation from .17 to .25 ($p < .001$). Longer reports touch more storylines mechanically, but high-uncertainty targets receive marginally shorter reports, so the extra breadth in high-uncertainty campaigns is more distinct storylines per unit of text, not merely more text.

The second row of Table D1 reports a discriminant check. If breadth were a response to market turbulence in general rather than to evaluative uncertainty specifically, the broad/focused split should load on pre-campaign volatility just as narrative share does. It does not: the same

comparison of means on volatility yields a gap of 0.00 standard deviations ($t = 0.00$, $p = .996$, $n = 780$), even though the identical split tracks evaluative uncertainty at 0.37 standard deviations.

The paper interprets the breadth–uncertainty association as short sellers adjusting their rhetoric to the presence or absence of shared evaluative standards. There are alternative readings. First, a report's storytelling tropes could in principle reflect what the report accuses the target of. We therefore coded each campaign on a second, independent content layer: ten campaign-level accusation domains (fraud, accounting manipulation, unsustainable business, governance red flags, stock promotion, and so on), kept deliberately separate from the nine segment-level narrative tropes. The two layers are substantively connected — broad-repertoire campaigns do allege more domains (a mean of 5.4 versus 3.5) — but they are not interchangeable: clustering campaigns on allegations alone fails to reproduce the broad/focused split (adjusted Rand index = .009 at $k = 2$, essentially chance agreement). More importantly, the accusation mix does not absorb the association. Partialling the number of alleged domains out of the breadth–uncertainty correlation leaves $r = .13$ ($p < .001$), and partialling out all ten domain indicators individually leaves $r = .09$ ($p = .015$), against .17 raw.

Second, if rhetorical breadth indicated more substantial information rather than a rhetorical choice, it should be associated with more complex target companies and with longer disclosures. Table D2 shows it is not. It reports the Pearson correlation between the broad-versus-focused indicator and each of eight standard proxies for target complexity and disclosure length; no measure correlates with breadth above $|r| = .09$, and the full bundle jointly explains $R^2 \approx .02$ of the variation in breadth. Nor does complexity account for the uncertainty association: partialling the well-covered complexity and disclosure measures (intangible and goodwill shares, 10-K

length, and 10-K complexity-word share) out of the breadth–uncertainty correlation leaves it unchanged at r = .17 (p < .001, n = 718), and adding the thinly covered segment-structure measures leaves r = .16 (p = .008) on the smaller sample they permit (n = 291).

The complexity and disclosure measures come from standard sources. Business-segment and geographic-segment counts are the number of segments the target reports in Compustat's segment files, and the segment-sales Herfindahl is the sum of squared segment sales shares, so that higher values indicate a business concentrated in fewer lines. The intangible-asset and goodwill shares scale each balance-sheet item by total assets, on the logic that intangible-heavy balance sheets are harder to value from public filings. Firm age is the listing age described in Appendix A. The two disclosure measures come from the target's most recent pre-campaign 10-K: length is the log word count, and the complexity-word share is the fraction of words on the Loughran–McDonald complexity list. All results in this Appendix are simple sample statistics. But every association can be re-estimated under nested control batteries (year, industry, leverage, and target fundamentals) and with standard errors clustered by short seller; no conclusion changes under any specification.

Table D1. Report Design and the Target Environment

Association	Statistic	Estimate	p	n
Content breadth (broad vs. focused) and evaluative uncertainty	Welch t-test on group means (t = 5.08)	+0.37 SD (r = .17)	< .001	809
Content breadth and volatility (discriminant)	Welch t-test on group means (t = 0.00)	+0.00 SD (r = .00)	.996	780
Narrative share and volatility, all reports	Pearson correlation	r = .13	< .001	772
… within broad reports	Pearson correlation	r = .24	< .001	548
… within focused reports	Pearson correlation	r = −.03	.621	224
Slope difference (content × narrative share)	Plain-OLS interaction (t = 4.20)	β = +1.12	< .001	772
Caveat: continuous trope count and evaluative uncertainty	Spearman correlation	ρ = .05	.135	809

Table D2. Report Breadth Is Not an Artifact of Target Complexity, Report Length, or Coverage Supply.

Predictor (Panel A) / Specification (Panel B)	Estimate	R^2	N
Panel A. Bivariate association of target-complexity and report-length measures with report breadth (Pearson r with the broad-repertoire indicator)			
Business-segment count	.01		521
Geographic-segment count	.00		344
Segment-sales Herfindahl	.00		465
Intangible-asset share	−.01		798
Goodwill share	−.00		801
Firm age (years since listing)	−.09		816
10-K length (log word count, LM)	.06		742
10-K complexity-word share (LM)	−.04		742
Full complexity + length bundle (joint model)		.02	291
Panel B. Evaluative-uncertainty/breadth associations (linear-probability coefficient on evaluative uncertainty)			
Evaluative uncertainty only	.55**	.03	809
+ volatility, year, industry	.47**	.10	779
+ report length (log words)	.60***	.36	779
+ visual-exhibit depth	.59***	.36	779
Alternative: + volatility, year, full short-seller fixed effects	.46***	.36	779

Note: Panel A reports Pearson correlations between the broad-versus-focused repertoire indicator and standard proxies for target complexity and disclosure length; N varies with the coverage of each source. Panel B regresses the broad-repertoire indicator on evaluative uncertainty. Rows 1–4 form the target-control and report-depth ladder; the final row is a separate within-seller specification with volatility and year controls plus a distinct fixed effect for every seller. Standard errors are clustered by seller throughout. Estimates are linear-probability coefficients on the 0/1 broad indicator. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table D3 reports the fit–return regression referenced in the findings. Fit is a tertile-corner construct with two legs: a content leg (a broad repertoire on a top-tertile-uncertainty target, or a focused one on a bottom-tertile target) and a register leg (top-tertile narrative share on a top-tertile-volatility target, or bottom-tertile narrative share on a bottom-tertile one); the joint indicator requires both legs to fit, and middle-tertile campaigns are neither fit nor counterfit. The dependent variable is the negative of the cumulative abnormal return over the stated window, so positive coefficients indicate larger abnormal declines. Coverage is the binding caveat: valid abnormal returns require CRSP matching, and the estimation sample (N = 351) over-represents larger, analyst-covered, lower-uncertainty targets — the excluded campaigns differ by up to 1.2 standard deviations of evaluative uncertainty — so the estimates corroborate the configurational argument

on the covered segment of the corpus rather than test it on the full sample. Under the unanimity vote rule the joint [0,+21] coefficient falls to +.043 (p = .32); Appendix E.7 discusses this sensitivity.

Table D3. Rhetorical Fit and One-Month Abnormal Returns (reinstated fit–return regression).

Variable	(1) Unadjusted	(2) Firm/year controls	(3) + Seller fixed effects	(4) + EUS & volatility
Panel A. Joint fit and one-month abnormal decline, [0,+21]				
Joint fit	.083*	.075*	.086*	.089*
	(.032)	(.031)	(.039)	(.043)
Log market capitalization		−.011	−.009	−.004
		(.008)	(.016)	(.017)
Book leverage		.090	−.051	−.166
		(.166)	(.171)	(.171)
Evaluative uncertainty				−.093
				(.217)
Pre-campaign volatility				.082
				(.087)
Announcement-year fixed effects	No	Yes	Yes	Yes
Two-digit NAICS fixed effects	No	Yes	Yes	Yes
Short-seller fixed effects	No	No	Yes	Yes
Direct EUS and volatility controls	No	No	No	Yes
Observations	351	351	351	351
Short-seller clusters	90	90	90	90
R^2	.014	.121	.382	.395

Panel B. Timing and component comparisons (fully adjusted)	**(5) Joint fit [0,+1]**	**(6) Joint fit [0,+21]**	**(7) Volatility–narrative fit [0,+21]**	**(8) EUS–content fit [0,+21]**
Fit	−.003	.089*	.132**	.030
	(.021)	(.043)	(.046)	(.037)
Log market capitalization	−.005	−.004	−.013	.007
	(.006)	(.017)	(.019)	(.015)
Book leverage	−.111	−.166	−.059	.005
	(.078)	(.171)	(.228)	(.106)
Evaluative uncertainty	.017	−.093	−.143	.050
	(.076)	(.217)	(.212)	(.134)
Pre-campaign volatility	.081*	.082	.078	.072
	(.034)	(.087)	(.077)	(.056)
Announcement-year fixed effects	Yes	Yes	Yes	Yes
Two-digit NAICS fixed effects	Yes	Yes	Yes	Yes
Short-seller fixed effects	Yes	Yes	Yes	Yes
Direct EUS and volatility controls	Yes	Yes	Yes	Yes

Variable	(1) Unadjusted	(2) Firm/year controls	(3) + Seller fixed effects	(4) + EUS & volatility
Observations	351	351	315	448
Short-seller clusters	90	90	81	97
R^2	.416	.395	.399	.324

Note: Entries are OLS coefficients with short-seller-clustered standard errors in parentheses. The dependent variable is abnormal decline (= −CAR); positive coefficients mean fit campaigns decline more. Fit is defined at matched tertile corners: EUS–content fit pairs broad reports with high EUS and focused reports with low EUS; volatility–narrative fit pairs narrative-heavy reports with high volatility and argument-heavy reports with low volatility. Joint fit requires both legs; either counterfit leg makes the joint campaign counterfit. Middle tertiles and missing cases are excluded. Panel A uses one complete-case sample: Model 2 adds firm controls and year/industry fixed effects, Model 3 adds seller fixed effects, and Model 4 adds EUS and volatility. Panel B uses Model 4 throughout; Model 6 repeats it as the comparison anchor. [0,+1] is the announcement window and [0,+21] the one-month drift. Fixed-effect coefficients and intercepts are omitted. Two-tailed tests. * $p < .05$, ** $p < .01$, *** $p < .001$. Two sensitivities bound the estimates: under the stricter unanimity vote rule the joint [0,+21] coefficient falls to +.043 ($p = .32$); and valid abnormal returns require CRSP coverage, so the estimation sample over-represents larger, analyst-covered, lower-uncertainty targets (excluded campaigns differ by up to 1.2 standard deviations of evaluative uncertainty). Read as corroboration on the covered segment of the corpus, not a full-sample test. Configurational associations, not causal estimates.

Appendix E — LLM Procedures and Reliability

This Appendix documents how we applied the frozen codebook to the full corpus using language-model coders, and the procedures by which we established that the resulting campaign-level constructs are reliable enough to support the claims made in the body of the paper. It covers (E.1) the construction of an adjudicated human ground-truth pool, (E.2) per-pool inter-coder reliability among the human team, (E.3) the LLM coding stack and its prompt protocol, (E.4) the fine-tuned audit voter, (E.5) segment-level model performance against the ground truth, (E.6) the measurement rules that translates segment-level outputs into campaign-level constructs, (E.7) the vote rules and the campaign-level reliability they deliver, and (E.8) held-out diagnostics on a testing pool reserved for stress-testing the stack.

E.1 The adjudicated human ground-truth pool

Validation of computational coding requires a reference standard against which model output can be evaluated. We constructed that standard ourselves rather than relying on a single-coder gold set, because precision and recall are interpretable only when human disagreement has been separated from model error. Two authors and two predoctoral research assistants dual-coded approximately 860 paragrpah-level segments across thirty-eight campaigns. Coding proceeded in four batches, with each pool dual-coded by a different subset of the four-coder team to expose coder-specific drift. A separate held-out pool of nine campaigns and one hundred eighty-five segments was coded by the two authors and never used during prompt refinement, training, or vote-rule selection; it serves as the corpus-side analogue of an out-of-sample test set and is reported in section E.8.

After each batch we adjudicated disagreements through structured discussion grounded in the codebook's inclusion and exclusion criteria. Any refinement that emerged from adjudication was logged in a coding-manual revision history, the codebook was updated, and prior pools were re-coded under the revised rule so that the final ground-truth pool reflects a single, consistent codebook rather than a moving target. After the fourth batch, the codebook was frozen. The freeze applies to category definitions only: at the case-comparison stage we develop new theoretical constructs freely, but we do not introduce new codes into the corpus measurement.

Once frozen, we used a set of LLM agents to apply the codebook to the corpus as a whole. We now discuss the reliability of the resulting codes. The discussion focuses on the constructs most relevant to the analysis: trope presence (the multi-label binary indicator of whether each of nine canonical blame-attribution patterns appears in a segment), tone (a three-level ordinal scale from professional through colorful to theatrical), narrative structure (a three-way nominal classification of segments as narrative, argumentative, or uncertain), and argument rigor (a three-level ordinal scale distinguishing unsupported claims, supported claims, and fully argued claims with explicit warrant). We report the reliability for allegations, a secondary construct used in Appendix D.

E.2 Per-pool inter-coder reliability

We report two layers of human reliability. The first is pairwise agreement and Cohen's κ before adjudication, which captures whether the codebook produces convergent judgments when applied by independent coders working in parallel. The second is post-adjudication reliability after structured discussion, which captures the reproducibility of the codebook as the corpus-wide measurement instrument. For ordinal constructs (tone and argument rigor) we use weighted κ

with linear weights so that adjacent-category disagreement is penalized less than extreme disagreement; for trope presence we report κ as the mean across the nine binary trope indicators.

Inter-coder reliability on the four headline constructs reached substantial-to-near-perfect agreement after the second batch: raw agreement of 99% on trope presence, 98% on tone, 97% on narrative structure, and 91% on argument rigor, corresponding to weighted Cohen's κ of approximately 0.93, 0.95, 0.93, and 0.84, respectively. The third and fourth batches sampled harder material — campaigns with denser overlapping tropes, more boundary-tone passages, and a heavier mix of narrative-versus-argument ambiguity — and per-pool agreement therefore dipped before adjudication. The adjudication protocol elevated post-resolution agreement across all four pools; the cross-pool mean Cohen's κ on trope presence in the final adjudicated pool is 0.84.

We dual-coded with adjudication rather than relying on a single-coder gold standard for two reasons. First, the codebook captures rhetorical strategies at high granularity — nine trope categories with overlapping family resemblances, three-level ordinal scales for tone and rigor — and category boundaries are exactly where measurement reliability is contested. Single-coder gold sets paper over those boundaries by definition. Second, language-model coders make characteristic errors that pattern with prompt design and provider-specific biases; without an adjudicated pool we could not separate prompt errors from genuine human ambiguity. The structured-disagreement protocol therefore did double duty: it stabilized the codebook and it produced the data we needed for downstream model evaluation.

E.3 The LLM Coding Process

We use a panel of three frontier API models plus a fine-tuned open-weight model as an audit voter. The three frontier models are Anthropic Claude Sonnet 4.6, OpenAI GPT-5.1, and

OpenAI GPT-5-mini. A fine-tuned Llama 3.3 70B variant (designated v7) trained on the adjudicated ground-truth pool serves as an independent audit voter for boundary categories where the API panel is known to over-fire. Each model receives task definitions based on the coding manual, including additional inclusion and exclusion criteria, prototypical examples, an output schema, and procedural instructions for ambiguous segments; temperature is set near zero, output is structured JSON validated against task-specific JSON schemas where the provider supports guided decoding, and each task is routed to its own prompt rather than being bundled. Tropes are coded conditionally on segments that the narrative-structure task classifies as narrative, so the trope question is asked only for segments where narrative is present.

We use a panel rather than a single best model because each model has known and idiosyncratic biases on rare or boundary categories. GPT-5.1 is the strongest ordinal coder; Claude Sonnet 4.6 is the strongest narrative-structure coder and has the lowest false-positive rate on conservative tropes; GPT-5-mini is a useful third voter because its error pattern is not collinear with either of the larger frontier models. The panel design lets us treat model disagreement as observable rather than averaged away: every segment carries the three API votes used by the primary rule; the Llama audit vote is complete for 822 of 830 campaigns and is retained as a separate sensitivity layer alongside the primary aggregate, so robustness can be tested by re-running aggregation under alternative vote rules without re-running the models.

The full corpus comprises 21,504 segment-level rows across 830 campaigns under coverage at the time of paper analysis. Of these, 21,490 segments cleared the Stage 01 substantive-content gate; 16,968 segments cleared the shared narrative gate that routes a segment into trope coding (a permissive gate triggered by either Claude or GPT-5.1 classifying the segment as narrative). Trope codes are therefore defined on 16,968 narrative segments per campaign across the

three API voters. Across these segments the three-API vote stack produced complete trope votes for all 16,968 segments; no segment required repair under the two-of-three valid-vote rule. The validation report attached to the final dataset documents these counts and the per-construct vote-state distributions.

Table E1. LLM agents and intended role.

Coder	Provider	Role	Vote position
Claude Sonnet 4.6	Anthropic	Frontier API	Primary voter (narrative-leading)
GPT-5.1	OpenAI	Frontier API	Primary voter (tone-leading)
GPT-5-mini	OpenAI	Frontier API	Primary voter (independence-leading)
Llama 3.3 70B v7 (LoRA-fine-tuned)	Self-hosted (HPC vLLM)	Fine-tuned audit voter	Audit / debiasing

Note: The three frontier API models form the panel that cast the primary votes. The fine-tuned Llama 3.3 70B model (v7) serves as an audit and debiasing signal on boundary categories and does not enter the primary vote.

E.4 The fine-tuned Llama audit voter

The fourth voter is a parameter-efficient fine-tune of Llama 3.3 70B Instruct using LoRA. The training set is the four adjudicated training pools — thirty-eight campaigns, approximately 2,800 adjudicated examples after task-conditional filtering. Held out from training, and never seen by the model at any point in the design loop, is a coding pool with nine campaigns that we use to evaluate performance. We do not treat the fine-tuned model as an equal fourth voter in the primary specification because the testing pool diagnostics show that Llama v7 is much more conservative than the API panel on positive trope identification. Substantively, the model is genuinely complementary to the API panel: its errors correlate with each API model at only $r \approx 0.28$ to 0.49, meaning that disagreements identify boundary cases the API panel is least independent about. It also correctly flips 48 to 62 percent of API primary-trope errors depending on which API model it is checked against. We therefore use it as an audit / debiasing signal.

E.5 Segment-level model performance against the ground truth

Applying a granular codebook reliably is demanding for any coder. Even at the second-batch peak, human agreement on trope presence reached 99% (weighted $\kappa \approx 0.93$) only after adjudication; before adjudication, agreement was substantially lower. Language-model coders inherit the same difficulty. We therefore report segment-level reliability against the adjudicated training pool not as evidence that any single model has reached human reliability. Rather, these numbers serve as evidence for which aggregation rule converts noisy segment-level vote vectors into stable campaign-level constructs. Table E2 reports the headline segment-level numbers.

Table E2. Segment-level model reliability against the adjudicated training pool for trope presence.

Specification	Macro F1	Cohen's κ	All-zero / NNT rate
Three-API majority	.525	.519	.036
Three-API unanimous	.520	.502	.117
Llama v7 production-gated (audit voter)	—	—	.062 (pool 13)
Single model (GPT-5.1, refresh)	.495	.414	—

Note: The all-zero / NNT rate is the share of trope-eligible segments receiving no positive trope vote; for the Llama v7 voter it is reported on held-out pool 13. Leading zeros are omitted for statistics bounded by 1.

Two patterns matter for interpretation. First, exact segment-level trope assignment remains noisy: current macro F1 is .525 under majority and .520 under unanimity. This is why the analysis aggregates segment level codes to campaign-level repertoires. Second, unanimity is not a missing-data solution: the same vote vectors are observed, but the stricter positive criterion raises precision while reducing recall and producing sparser campaign repertoires. Majority is therefore the coverage-preserving primary construction; unanimity is a conservative sensitivity check.

E.6 The measurement contract: from segments to campaigns

Because segment-level labels are too noisy to be direct outcomes, we impose measurement rules that tie every campaign-level construct used in the analysis to a documented aggregation rule and a documented reliability target. We use two rules: first, the legitimate unit of analysis is the campaign, not the segment. Segment-level labels are intermediate computational artifacts and never appear in the analysis as predictors or outcomes. Every campaign-level construct is defined as an aggregation of segment-level votes. Second, the empirical analysis uses the campaign constructs only for analyses that their reliability supports. For instance, the trope-repertoire construct is stable enough under the majority rule to do headline work in the clustering analysis. But it abstracts substantially from the segment-level codes: it merely records whether or not a trope is present anywhere in the campaign. Similarly, tone is measured as a campaign-level average because assignments of the most extreme option (theatrical) are noisy on the segment level. Tone-average is the mean of the segment-level ordinal tone code over substantive segments (1 = professional, 2 = colorful, 3 = theatrical). Narrative share is the share of substantive segments classified as narrative by the resolved three-API majority rule. Argument-rigor average is the analogous mean on the rigor scale (1 = unsupported, 2 = supported, 3 = argued).

E.7 Vote rules and campaign-level reliability

Segment-level vote vectors are aggregated under explicit vote rules whose reliability against the adjudicated training pool is audited per construct. The primary specification is a three-API majority rule: a segment is coded positive for a label if at least two of the three frontier models code it positive. The unanimous rule, under which all three models must agree, serves as a precision-oriented robustness specification; it sharply reduces recall and compresses campaign repertoires, so it is not required to reproduce every headline model. Three four-voter variants that

incorporate the Llama audit voter and an any-vote rule serve as targeted stress tests. Where an alternative rule changes a substantive interpretation, we report that sensitivity explicitly.

Table E.3 reports the full audit for the analysis constructs under the primary three-API majority rule; the unanimous rule is audited at the segment level in Table E2. For each construct, the table records where the construct is coded and what it is aggregated to, how well the three models agree with one another at both levels, and how well the aggregated value matches the adjudicated ground truth. Because the models' errors are correlated within a document rather than independent across segments, we compute the campaign-level agreement coefficients directly instead of projecting them from segment-level agreement.

Two features of the audit matter for interpretation. First, the campaign-level coefficients are absolute-agreement measures — nominal α for the binary constructs, ICC(2,1) for the continuous ones — so a construct scores low when the models rank campaigns alike but disagree on the level. Narrative share is the one row where that penalty dominates: the three models' campaign values correlate at a mean pairwise r of .65, so their disagreement concerns the level, not the ordering. Second, the models over-detect on every content construct (on trope presence, recall is .844 against a precision of .641), which is why the bias terms are positive there. Scaled by the ground-truth standard deviation, the narrative over-call is the largest distortion in the measurement chain and trope breadth the second largest, while tone and rigor are small. Neither over-call bears on a headline claim. Narrative share enters the analysis as tertiles, and removing the constant shift raises its ICC to .52. Trope breadth enters only through a cut that is data-driven on the model's own distribution, and the models preserve the ranking that cut depends on (Spearman $\rho = .70$).

Table E3. Full Measurement Audit of the Analysis Constructs: Coding Base, Inter-Model Agreement, and Validity against Ground Truth.

		Coding decisions		*Inter-model reliability (no ground truth)*		*Validity vs. adjudicated ground truth*					
Construct	**Coding unit**	**GT**	**Base**	**Segment**	**Campaign**	**Macro F1**	**κ**	**r**	**ICC(2,1)**	**Bias**	**Bias (SD)**
Trope presence, binary (9 tropes, pooled)	Segment × trope → campaign	3,519	152,712	.670	.693	.717	.569	—	—	+0.13	+0.27
Trope breadth (no. of trope families)	Segment × trope → campaign	3,519	152,712	.670	.826	—	—	.724	.477	+1.84	+0.91
Allegations (10 domains, pooled)	Campaign × domain	370	8,300	—	.786	.746	.500	—	—	+0.13	+0.27
Narrative share (narrative_pct)	Segment → campaign	787	21,490	.414	.372	—	—	.555	.274	+0.28	+1.13
Tone (tone_avg)	Segment → campaign	788	21,490	.442	.760	—	—	.801	.615	−0.14	−0.35
Argument rigor (argument_rigor_avg)	Segment → campaign	600	18,331	.495	.615	—	—	.745	.747	−0.03	−0.09

Note: The full measurement audit behind Table 4 in the paper, under the production specification — three-API (Claude Sonnet 4.6, GPT-5.1, GPT-5-mini) majority vote — against the adjudicated ground-truth pool of ≈ 38 campaigns. GT counts the adjudicated decisions underlying each row; Base counts the aligned model decisions on which the inter-model coefficients are computed (21,504 segments across 830 campaigns, each coded by all three models). Inter-model reliability is Krippendorff's α on the raw segment decision and, at the campaign level, nominal α (binary constructs) or ICC(2,1) (continuous constructs) across the three models. Validity against ground truth is macro F1 and Cohen's κ for the categorical constructs and Pearson r and ICC(2,1) for the continuous ones; Bias is the model − ground-truth shift, reported in raw units and in ground-truth standard deviations. Argument rigor's GT count is below its 768 adjudicated segments because segments adjudicated as other or summary carry no rating. Genre and coherence are omitted because the analysis does not use them. Leading zeros are omitted for statistics bounded by 1.

Appendix F — Case-Level Illustrations of the Three Report Strategies

This Appendix reproduces, in full, the case table condensed in the main text. Each row summarizes a campaign, the strategy it uses, the evaluative environment of its target, whether that pairing fits the contingency theory, and the cumulative abnormal return of the campaign. A campaign counts as successful with it had negative cumulative abnormal return.

Table F1. Instances of the Three Report Strategies across Target Environments.

Case	Strategy	EUS · Vol	Fit	How the campaign works and why it fits	Outcome
Low EUS — fitting strategy: Focused					
Motorola Solutions Citron Research · 2017 ✓	Focused	low · low	Fit	Citron makes one argument, that the company's near-monopoly pricing on first-responder radios is about to meet competition, and dramatizes it with a single example of a radio sold for several times its United Kingdom price.	Success (-3%, -0.75σ)
Chegg Citron Research · 2019 ✓	Focused	low · low	Fit	The report stakes everything on one reframe, that the ed-tech company is really 'the poster child for institutionalized academic cheating,' timed to the Rick Singer admissions scandal that had just made academic integrity salient.	Success (-11%, -1.22σ)
Transdigm Group Citron Research · 2017 ✓	Focused	low · low	Fit	Compresses the entire case into a single analogy — the aerospace-parts maker as 'the Valeant of the Aerospace Industry' — pegged to the incoming administration's promise to cut military procurement, which would expose the firm's reliance on serial price hikes.	Success (-4%, -0.53σ)
Axos Financial Real Talk Investments · 2015 ✓	Focused	low · high	Fit	Develops one narrative of corporate obfuscation: that the transcript the bank filed as an 8-K exhibit is 'materially different than the BOFI Transcript' of its own investor call and of independent renderings, setting the versions side by side and declining the broader whistleblower allegations to keep the reader on one checkable object.	Success (-24%, -0.88σ)
AECOM Spruce Point Capital · 2016 ✓	Focused	low · low	Fit	Spruce Point rests the whole report on one forensic claim, that the engineering roll-up's post-merger free cash flow is 'overstated by approximately 90%,' then re-runs peer multiples on the corrected numbers to reach its target.	Success (-19%, -2.08σ)
NVIDIA Corporation Citron Research · 2017 ✓	Focused	low · low	Fit	This short 650-word note simply reclassifies a momentum favorite as a gamble rather than an investment, 'the moment that separates the gamblers from the investors,' alleging no wrongdoing and disputing only the valuation.	Success (-12%, -0.83σ)
Ideanomics Hindenburg Research · 2020 ✓	Focused	low · high	Fit	A theatrical two-part Twitter thread calls the company 'an egregious and obvious fraud,' compressing a pump-and-dump history of six chief executives and four corporate names, plus side-by-side photos, into one fast verdict.	Success (-66%, -1.00σ)
Applied Digital Corporation The Friendly Bear · 2023 ✓	Focused	low · high	Fit	Builds a single argument in a tight legal-forensic register, focused on ‘hidden skeleton’ tropes: the AI-datacenter highflyer is secretly a controlled company of its patron investment bank, walking through director-independence tests, a conveniently timed related-party loan, and governance red flags.	Success (-69%, -1.31σ)

Case	Strategy	EUS · Vol	Fit	How the campaign works and why it fits	Outcome
Bread Financial Holdings Citron Research · 2016 ✓	Broad · narrative	low · low	Misfit	Citron reframes the credit-card lender as 'a bank in the Halloween costume of a business services company' and, riding a fresh sell-side note, predicts Wall Street will soon mark it down. Uptake was shallow and fleeting: a same-day repost and a brief, automated-looking price dip that reversed within hours, after which attention turned to an activist's new stake and the stock recovered.	No move (+9%, +1.10σ)
Axos Financial The Friendly Bear · 2015 ✓	Broad · narrative	low · low	Misfit	Follows a New York Times investigation by arguing the branchless bank's peer-beating margins rest on risky mortgages to 'unsavory' borrowers — then layers further narratives of compliance failures, undisclosed self-dealing, and weak regulatory standing around that core claim. The bank never named the seller publicly but subpoenaed 'The Friendly Bear' in litigation; the announcement dip reversed and the stock recovered within the month.	No move (+4%, +0.50σ)
Burlington Stores Spruce Point Capital · 2016 ✓	Broad · argument	low · low	Misfit	Stacks a self-described 'plethora' of accounting red flags across a fifty-nine-slide deck, each headed 'Warning:', with no single story or argument dominating the presentation. Bloomberg and Benzinga ran the story and the company denied it in an 8-K filing, but no analyst engaged and the announcement dip fully reversed.	No move (+13%, +2.29σ)
Advanced Micro Devices Viceroy Research · 2018	Broad · argument	low · low	Misfit	Viceroy amplifies a third-party security firm's findings of 'severe and pervasive security flaws' in AMD's chips to argue the processors are 'unpurchaseable and outright dangerous', that 'the AMD growth story is dead', and that the company is 'worth $0.00' and must file for bankruptcy. The stock did not respond materially.	No move (-2%, -0.17σ)
Align Technology Spruce Point Capital · 2020	Broad · argument	low · high	Misfit	Spruce Point argues point by point that patent expirations and cheap 3D-printing labs are turning the company's clear aligners into a commodity, using term sheets, orthodontist surveys, and price cuts to show margins about to collapse. Investors did not react and the stock drifted higher.	No move (+7%, +0.23σ)
Bristow Group MOX Reports · 2016	Broad · narrative	low · high	Misfit	MOX Reports argues that the debt-laden helicopter operator's 'massive unhedged currency exposures' to the collapsing pound and Nigerian naira, still undisclosed to investors, will surface as losses on the coming earnings call, the 'near term catalyst' for a steep drop toward default. The stock did not move materially in the weeks that followed.	No move (-0%, -0.00σ)
MSCI Spruce Point Capital · 2024	Broad · argument	low · low	Misfit	Spruce Point recasts the index provider as 'a struggling company failing to transform itself', propping up earnings through 'value-destructive' and 'nepotism-like' acquisitions that benefit its former parent and alumni and through 'abusive financial reporting and accounting tactics', pegging 55 to 65 percent downside. The stock did not respond materially.	No move (-0%, -0.00σ)
Stryker Corporation Spruce Point Capital · 2022	Broad · argument	low · low	Misfit	Spruce Point's forensic report casts the device maker as a 'debt-fueled roll-up of medical product companies' that 'systematically overpays for acquisitions' and is 'covering-up inventory accounting problems' amid a margin squeeze, projecting 35 to 75 percent downside. The stock did not respond materially.	No move (-1%, -0.17σ)
High EUS · quiet — fitting strategy: Broad · argument					
Endurance International Group Gotham City Research LLC · 2015 ✓	Broad · argument	high · low	Fit	Assembles several separable arguments and narratives — revenue quality read from the roll-up's own disclosures against irreconcilable Indian regulatory filings, subsidiary fraud, reputational and regulatory exposure from its customers, and an unsustainable debt load — each resting on its own evidence rather than a single thesis.	Success (-8%, -0.65σ)

Case	Strategy	EUS · Vol	Fit	How the campaign works and why it fits	Outcome
PetIQ Spruce Point Capital · 2019 ✓	Broad · argument	high · low	Fit	Combines a genealogical exposé of the executives, grey-market drug diversion, rebate-dependent earnings, a cash-burning clinic rollout, and a plain peer valuation. Each argument or narrative offers an independent reason to short.	Success (-7%, -0.59σ)
Weis Markets Spruce Point Capital · 2018	Broad · argument	high · low	Fit	Stacks separable forensic flags — its own price checks and credit-card-panel data, auditor turnover, and an accounting choice that counts 'funds in transit under 7 days' as cash — to argue a grocer is masking financial strain.	Success (-8%, -1.23σ)
Tootsie Roll Industries Spruce Point Capital · 2017	Broad · argument	high · low	Fit	Spruce Point builds a broad forensic case against an 'inferior candy confectioner,' citing declining brands, a roughly 800-basis-point gross-margin overstatement, 'cookie jar' cash-flow inflation, and safety findings from public records, all under entrenched family control.	Success (-1%, -0.27σ)
Runway Growth Finance Corp NINGI Research · 2023 ✓	Broad · narrative	high · low	Misfit	Several separable lines of argument (understated loan values, dividends paid out of capital, and marked-up asset values at a business-development lender), but heavily relies on narratives about distress at different portfolio companies to tie the report together. The market barely responded and the stock was roughly flat over the following month.	No move (+5%, +0.66σ)
Trupanion Long Short Value · 2017	Broad · narrative	high · low	Misfit	Narrates a focused overvaluation story around the pet insurer's 'razor thin insurance margins' and 'no real moat outside of brand name.' If margins really are thin the growth story unwinds, and if they are fat, competitors will flood a crowding market. The market did not respond materially over the following weeks.	No move (-2%, -0.24σ)
Everyday Health Cliffside Research · 2016	Broad · narrative	high · low	Misfit	Cliffside recasts the health-media company as a 'poor mans WebMD' whose post-IPO acquisition spree disguises a declining core business and 'poor earnings quality,' arguing management leans on adjusted EBITDA to hide stalling organic growth. The market did not move materially over the following month.	No move (-0%, -0.01σ)
PennyMac Financial Services FFJ · 2022	Focused	high · low	Misfit	FFJ warns that the mortgage lender quietly built a large, undisclosed book of VA loans that the government only partly guarantees, so that rising rates and defaults could wipe out its thin equity. Investors were unpersuaded and the shares rose over the next month.	No move (+14%, +1.12σ)
Pershing Gold Hindenburg Research · 2017	Broad · narrative	high · low	Misfit	Hindenburg strings together the controlling insider's pump-and-dump associations, fraud-tainted early backers, and a mine that has 'produced zero mining revenue since inception' to argue the shares are nearly worthless. The market did not follow and the stock was little changed.	No move (+7%, +0.90σ)
National Beverage Glaucus Research Group · 2016	Focused	high · low	Misfit	Glaucus reads the sparkling-water maker's suspiciously high operating margins as manipulated earnings, citing a former attorney's sworn allegation that the chief executive admitted hitting targets through a 'little jewel box' and fabricated 'fake invoices,' all shielded by off-balance-sheet management entities. The market did not respond materially over the following weeks.	No move (-2%, -0.22σ)
Everyday Health The Street Sweeper · 2016	Focused	high · low	Misfit	The Street Sweeper argues, in a heavy medical metaphor, that a 'WebMD wannabe' wholly dependent on Google search is being cut out by Google's own symptom results and by ad-blockers. The report gained no traction and the stock climbed over the following month.	No move (+31%, +3.01σ)
MediciNova Mako Research · 2016	Broad · narrative	high · low	Misfit	Mako mocks a serial fundraiser whose three aging drug compounds have never cleared a late-stage trial, juxtaposing near-identical pipeline slides from 2007, 2009, and 2016 and foregrounding an undisclosed kickback lawsuit against the chief executive. The attack never left Seeking Alpha's comment section, drawing no media, analyst, or law-firm	No move (+36%, +2.50σ)

Case	Strategy	EUS · Vol	Fit	How the campaign works and why it fits	Outcome
				pickup, while the company stayed publicly silent and positive news pushed the stock up about 36%.	
High EUS · activated — fitting strategy: Broad · narrative					
Clearside Biomedical Cliffside Research · 2016 ✓	Broad · narrative	high · high	Fit	Concedes the bull case up front — the biotech is 'a fine company' likely to win approval — then pulls a re-derived patient count, a cautionary drug history, and insider red flags into a single chronological arc that ends with a clear prediction: insiders already own the good news and will exit when shares come off lock-up and early sales disappoint.	Success (-90%, -2.57σ)
Nikola Hindenburg Research · 2020 ✓	Broad · narrative	high · high	Fit	Threads a wide set of revelations through the founder's biography — recorded calls, texts, and a video of a truck rolling downhill — into one overarching character narrative of lifelong deception that ends by asking whether he is a visionary or 'a con artist.'	Success (-62%, -1.34σ)
Mullen Automotive Hindenburg Research · 2022 ✓	Broad · narrative	high · high	Fit	Hindenburg frames the EV startup generically as 'yet another fast talking EV hustle,' puncturing each promotional claim with import records, regulatory filings, and interviews in which supposed partners deny the touted deals ever existed.	Success (-45%, -0.54σ)
Intelligent Systems Marcus Aurelius Value · 2019 ✓	Broad · narrative	high · high	Fit	Fuses a governance exposé of a named director — the disgraced former chief executive of MiMedx installed as the audit committee's financial expert — with a re-model of the fintech as 'largely a niche Indian outsourcing business,' binding the two lines into a single story of concealment.	Success (-56%, -1.96σ)
LuxUrban Hotels Bleecker Street Research · 2024	Broad · narrative	high · high	Fit	Bleecker Street shorts the hotel-leasing company as 'The WeWork of Hotels,' arguing its stunning reported revenue hides undisclosed lawsuits over unpaid rent that could 'blow a hole in its balance sheet,' hotel deals announced but never signed, and published room rates that do not match actual booking prices. The stock fell sharply over the following weeks.	Success (-57%, -2.28σ)
Allakos Seligman Investments · 2019	Broad · narrative	high · high	Fit	Seligman weaves testimony from a trial investigator, a physician, and a patient's parent into a fraud warning, calling the biotech's key study a 'Phase 2 farce' run by a company that served as its own clinical-research organization.	Success (-53%, -1.92σ)
IonQ Scorpion Capital · 2022	Broad · narrative	high · high	Fit	Scorpion assembles 25 insider and expert interviews to recast the company's flagship 'world's most powerful quantum computer' as 'a useless toy' and a 'brazen hoax,' alleging staged photos, fictitious revenue, and fabricated executive credentials.	Success (-28%, -1.04σ)
Control4 Cannell Capital · 2016 ✓	Broad · narrative	mid · mid	Fit (boundary)	Ties a variegated set of complaints — financial problems, unfavorable product comparisons and reviews, insider share sales, board resignations, potential stock hype — together with an obfuscation narrative, concluding the company 'hides behind adjustments and acquisitions to obscure' its lack of profits. A mid-uncertainty boundary case, excluded from the theory tests; the stock rose over the following month.	No move (+19%, +0.91σ)
Digital World Acquisition Kerrisdale Capital · 2022	Broad · narrative	high · high	Fit	Kerrisdale tells the story of a Trump-media SPAC through its 'cast of characters' and their disregard for securities rules, walking a dated chronology of red flags to conclude the merger will never close.	Success (-63%, -2.52σ)
Comstock Mining Spruce Point Capital · 2021 ✓	Broad · argument	high · high	Misfit	Spruce Point fuses two micro-caps into one thesis as 'egregious pretenders' riding electric-vehicle-recycling hype, stacking red-flag tables under prosecutorial questions and valuing the equity 'at a multiple of its cash.' The report was received - a three-hundred-person	No move (-13%, -0.15σ)

Case	Strategy	EUS · Vol	Fit	How the campaign works and why it fits	Outcome
				shareholder call fielded a question about it the next day - but drew no rebuttal, no analyst, and no price move, even as its co-target fought back publicly.	
Faraday Future Intelligent Electric J Capital Research · 2021	Broad · argument	high · high	Misfit	J Capital argues on several fronts, a fraud-tainted founder, unexplained cash burn, five abandoned factories, and faked reservations, that the EV SPAC will never 'sell a car,' backing the case with site visits and former-employee interviews. Investors did not follow and the shares rose modestly over the next month.	No move (+11%, +0.40σ)
Purecycle Technologies Holdings Hindenburg Research · 2021	Broad · argument	high · high	Misfit	Attacks the plastics-recycling SPAC on several argumentative fronts — 'the latest zero-revenue, ESG-themed SPAC,' executives with a record of 'destroying public shareholder capital,' and a licensed process a polymer expert calls 'a bomb' — without binding them under a single overarching narrative. The market did not move.	No move (-6%, -0.23σ)
IVERIC bio Utopia Capital Research · 2019	Focused	high · high	Misfit	Utopia argues by analogy that the rebranded eye-disease company is repeating its old playbook, hyping a mid-stage trial result and then diluting into it, after earlier trials failed. Investors were unpersuaded and the stock rose over the following month.	No move (+14%, +0.26σ)
Proteostasis Therapeutics Utopia Capital Research · 2019	Focused	high · high	Misfit	Utopia argues the biotech's 300 percent run is unjustified because its lead drug does not work, leaning on an earlier report that its trial's placebo group was 'anomalously bad,' then adding dilution risk and a questionable shareholder. The market was unmoved and the stock ended roughly flat.	No move (+2%, +0.05σ)
Synthesis Energy Systems White Diamond · 2019	Focused	high · high	Misfit	White Diamond argues the coal-gasification company's 'failed clean energy technology' produces no revenue, making it 'essentially an empty shell with worthless assets' headed for a reverse-split collapse. The market did not respond materially.	No move (-4%, -0.13σ)
Polarityte Cliffside Research · 2017	Broad · argument	high · high	Misfit	Cliffside recasts a hyped 'regenerative medicine' reverse-merger as a diluted shell, arguing its real market value is closer to $300 million than the stated $83 million and that its lead product has been tested only on animals. The market shrugged and the stock rose over the following month.	No move (+27%, +0.80σ)
R1 RCM Jehoshaphat Research · 2023 ✓	Broad · narrative	high · low	Misfit	Jehoshaphat's 95-page report prosecutes a two-sided accounting indictment of the healthcare revenue-cycle manager: revenue rests on subjective 'mark-to-model accounting' while deferred expenses leave adjusted EBITDA 'inflated by approximately 55%' — 'vapor EBITDA.' The register misfits the quiet environment, yet the stock fell over the drift window; the close read attributes the move to three dated near-term catalysts bolted onto the muckraker frame, with earnings inside the window.	Decline (-23%, -1.85σ)

Note: N = 44 cases that illustrate the theory (up to 8 per cell), showing the recurrence of each report strategy in the environment where the design theory predicts it. Cases are grouped by target environment (bold rows); within each, Fit lists the fitting strategy that produced a decline (confirming fit) and Misfit lists mismatched strategies that did not (confirming misfit). A check (✓) marks the 21 focal cases carrying a verified close read and audience-reaction record in the comparison-case design. EUS and Vol are the campaign's evaluative-uncertainty and 60-day pre-campaign volatility tertiles. Outcome reports the post-report drift CAR [0,+21] and whether the target declined; 'Success' = a drift of at least 0.25 standard deviations of the target's own 21-day return distribution (shown after the raw CAR), so that a small raw move on a highly volatile stock is scored as 'No move' rather than a decline