Financial Influencers and Mispricing*

Yihong Huang Darcy Pu Ziqing Zhao
Guanghua School of Management, Peking University

March 20, 2026

Abstract

We develop a tractable model of financial influencers ("finfluencers") as strategic information intermediaries. A finfluencer observes fundamentals, chooses a message and a fee, and sells access to retail investors who prefer belief-confirming advice. Because investors decide whether to subscribe and trade, communication affects both beliefs and market participation. This feedback allows a profit-maximising finfluencer to shade bad news towards bullish recommendations, generating overvaluation in weak states, excess volatility, and predictable reversals when payoffs are realised. The model also delivers an accountability gradient: distortions are larger when reputational or platform discipline is weaker.

Keywords: financial influencers; finfluencers; social media; confirmation bias; endogenous participation; information design; mispricing; limits to arbitrage.
JEL codes: G12, G14, D82, D83, L82.

*This draft is preliminary. Emails: Yihong Huang (yihongh@gsm.pku.edu.cn), Darcy Pu (darcypu@pku.edu.cn, corresponding author), and Ziqing Zhao (ziqing.zhao@stu.pku.edu.cn). We thank James Dow, Linlin Ma, and Yan Xiong for helpful comments and suggestions. All errors are our own.

1 Introduction

A defining feature of modern retail trading is not merely that information travels fast, but that it is sold and consumed in *tailored* form. Financial influencers—"finfluencers"—on platforms such as Instagram, Weibo, X (formerly Twitter), Telegram, YouTube, and forum-based communities (e.g., WallStreetBets) monetise attention through subscriptions, paid groups, affiliate links, and engagement-driven advertising (Cookson et al., 2024c; Pedersen, 2022). Unlike traditional mass media, which largely monetises broad reach (Mullainathan and Shleifer, 2005), a finfluencer's revenue scales with conversion and retention: a per-capita monetisation rate times an endogenously formed subscriber base. This seemingly small institutional difference changes the economics of communication in financial markets—and, as we show, can generate systematic mispricing.

Why does this matter for prices? In a market with Bayesian retail investors, informed traders, and noise traders, one might expect biased communication to wash out in equilibrium: investors can discount a sender with commercial incentives, and standard noisy rational-expectations logic often delivers extra volatility rather than systematic price bias (Grossman and Stiglitz, 1980; Hellwig, 1980). Our paper shows that this intuition can fail once communication changes who enters the market and therefore what information prices aggregate. A key distinction is that the uninformed retail investors targeted by the finfluencer are not noise traders: they hold heterogeneous prior beliefs about fundamentals, update rationally on the finfluencer's report and the equilibrium price, and make endogenous subscription and trading decisions. Their confirmation motives affect both subscription and trading participation, while the finfluencer chooses content and fees anticipating those participation responses. The model is needed to trace that equilibrium feedback: slanted communication reshapes the trading population; the trading population reshapes price informativeness; and the resulting price system feeds back into the value of slant itself. Under those conditions, even fully Bayesian subscribers need not undo the distortion, and equilibrium prices can become systematically biased.

Three institutional features distinguish finfluencers from the traditional financial media studied by Goldman et al. (2022) and the mass-media outlets analysed by Mullainathan and Shleifer (2005), and motivate our modelling choices. *First, niche audiences and extreme selection.* Mass media outlets speak to broad, heterogeneous audiences. Finfluencers typically attract a narrow, self-selected clientele—often more extreme in tastes and beliefs—because following is voluntary, low-friction, and algorithmically reinforced (Cookson et al., 2023). This selection makes belief heterogeneity and confirmation incentives first-order concerns. *Second,*

interactivity and repeated reinforcement. Finfluencer communication is high-frequency and interactive: comments, live streams, private groups, repost chains, and "community" norms. The investor consumes not only information but also identity and social validation, which naturally maps to belief-dependent utility costs of disconfirmation. *Third, platform algorithms reward engagement, not accuracy.* Distribution is mediated by ranking and recommendation systems optimised for engagement (Allcott et al., 2022). Content that triggers emotion, identity, or group cohesion can be amplified even if accuracy is low, complementing the finfluencer's revenue motive: engagement expands the subscriber base and raises advertising rates. A further consequence is that finfluencer content can be *publicly observable yet privately processed*: the sheer number of finfluencers and the volume of content mean no individual investor can feasibly monitor and aggregate all messages. In practice, each retail investor follows a small set of accounts, so finfluencer signals function like *club goods* or "quasi-private" information.

Precisely because of these institutional features, the channel through which finfluencers affect asset prices differs fundamentally from that of traditional media. In the mass-media framework of Mullainathan and Shleifer (2005), competing newspapers slant toward the prior beliefs of a broad readership, but prices are not modelled. In the financial-journalism model of Goldman et al. (2022), a benevolent journalist selects which corporate announcements to report and attempts to debias managerial obfuscation; the journalist speaks to a broad readership who take articles at face value. Our finfluencer, by contrast, is a *revenue-maximising* intermediary who observes fundamentals and designs both the *content* and the *price* of information for an endogenously self-selected niche of subscribers whose willingness to pay depends on belief-confirmation motives. The model's central contribution is a feedback loop linking content design, audience composition, and price informativeness: the finfluencer's slant determines who subscribes; subscribers' trades determine price informativeness; and anticipated informativeness shapes the finfluencer's ex ante content choice.

We build a tractable four-date model with a risky asset in zero net supply. A binary fundamental $\theta \in \{L, H\}$ determines the terminal dividend d_θ at $t = 3$. At $t = 1$, a finfluencer perfectly observes θ and chooses: (i) a discrete report $s_I \in \{L, H\}$, and (ii) a subscription fee c_F (or, more generally, a per-subscriber monetisation rate capturing subscriptions plus advertising). At $t = 2$, a mass χ of uninformed retail investors each has a private leaning (type) that shapes priors and generates a psychological cost ϕ when the finfluencer's message contradicts their leaning. Investors choose whether to subscribe; only subscribers trade.[1]

[1]This assumption is a tractable reduced form for participation frictions and adverse-selection concerns. It is shared by Goldman et al. (2022), who make the analogous assumption that investors who do not read the financial newspaper do not trade. We show in Section E.1 that our results survive when non-subscribers

Informed traders observe θ and trade competitively; liquidity demand is noisy. Quadratic trading costs prevent infinite positions and capture limits to arbitrage. A convex manipulation cost $\frac{\mu}{2}m^2$ disciplines the finfluencer's choice of slant. We interpret μ as reputational capital, verification intensity, platform oversight, and the cumulative effort required to sustain increasingly aggressive misreporting. Low μ corresponds to the lightly supervised influencer environment; high μ corresponds to traditional media, regulated research, or high-credibility outlets. The comparison between these intermediaries therefore turns not on different "tastes" for truth, but on different levels of credibility at stake, in the spirit of Gentzkow and Shapiro (2006). Figure 1 provides a flow chart of the timeline, the decisions available to each agent, and the potential payoffs.

A natural objection is that if investors know the finfluencer's incentives and strategy, they can "debias" the report. In a continuous-signal environment with an additive bias, investors who rationally anticipate the expected bias may indeed net it out. Our mechanism is fundamentally different for three reasons. First, the finfluencer's report is *discrete and strategic*: without loss of generality, when there are more optimistic investors than pessimists, she pools states by reporting H with probability m even when $\theta = L$. This is a state-contingent pooling rule, not an additive shift that can be subtracted. Second, liquidity noise makes equilibrium prices *imperfectly revealing*. Even after Bayesian updating on the report and the price, subscribers typically cannot infer θ with certainty. Third, *endogenous participation distorts aggregation*: the set of traders in the market depends on the report and fee. Even if each subscriber updates rationally, the price reflects a selected population's demand and the finfluencer's induced information environment. We formalise and contrast these mechanisms in Section F, where we show that in a continuous additive-slant benchmark, the bias is fully anticipated and expected mispricing is zero.

The model delivers three main implications. *(i) Endogenous mispricing from strategic pooling.* Mispricing arises endogenously from the interaction between confirmation-driven participation and monetisation incentives: when the influencer pools bad fundamentals with bullish communication, prices become systematically biased upward in low-fundamental states (Propositions 8 and 9). *(ii) Reputational discipline and media type.* Manipulation is increasing in confirmation bias and decreasing in the manipulation-cost parameter μ, implying that low-accountability finfluencers manipulate more than traditional media with larger reputational stakes (Proposition 6). The zero-discipline benchmark is bang-bang; the empirically richer interior case requires positive discipline strong enough to regularise the Strategy 2 objective. A direct empirical implication is a platform-accountability gradient: mispricing should be

trade with attenuated intensity.

strongest on low-μ venues such as pseudonymous chat groups or lightly moderated short-video platforms, and weaker on high-μ venues with track records, verification, or institutional oversight. *(iii) Predictable reversals and welfare consequences.* Because overvaluation stems from strategic pooling and noisy aggregation, the model predicts return reversals when fundamentals are revealed, alongside welfare losses from costly subscriptions and distorted information production (Proposition 10).

Our key results are robust to several extensions: allowing non-subscribers to trade with attenuated intensity, endogenising reputational discipline through repeated interactions, modelling competition among multiple finfluencers, and treating algorithmic amplification as a primitive. We discuss these in detail in Section E, noting here that the central mechanism—endogenous subscription selection driven by confirmation preferences feeding back into finfluencer content design—survives in all variants. What changes is the *magnitude*: stronger reputational discipline compresses $m^\star$ but does not eliminate it when ϕ is large enough; competition among finfluencers can either discipline or amplify slant depending on whether differentiation or verification forces dominate.[2]

The remainder of the paper is organised as follows. Section 2 situates our mechanism in the literatures on financial media, persuasion and slant, social transmission, and limits to arbitrage. Section 3 lays out the model and equilibrium concept. Section 4 solves for equilibrium pricing and characterises when prices are (imperfectly) revealing. Section 5 analyses endogenous subscription, the finfluencer's optimal fee and reporting policy, and how audience selection feeds back into price informativeness. Section 6 derives the model's mispricing implications and comparative statics. Section 7 studies welfare and discusses policy levers. Section 8 concludes.

2 Related Literature

Our paper sits at the intersection of financial media theory, information design, behavioural finance, and rational-expectations asset pricing. We organise the discussion around four themes: intermediary-driven information production, media slant and persuasion, social influence and echo chambers, and the empirical literature on social media and retail trading.

The closest theoretical antecedent is Goldman et al. (2022), who model a financial

[2]Intuitively, competition disciplines slant when subscribers can verify finfluencer accuracy through alternative sources, but amplifies slant when competing finfluencers differentiate by targeting increasingly extreme belief niches—much as competing newspapers in Mullainathan and Shleifer (2005) can produce more slant than a monopolist.

journalist that selects which corporate announcements to report and attempts to debias managerial obfuscation. In their framework the journalist is *benevolent*: she reports to help readers trade profitably, and investors who do not read the newspaper do not trade. Because the journalist tries to *remove* bias, residual mispricing is a by-product of the firm manager's strategic obfuscation, not of the journalist's own incentives. Our setting reverses the fundamental source of slant. The finfluencer is a *revenue-maximising* agent who observes fundamentals and strategically designs both the content and the price of information to cater to a self-selected niche audience. Slant is an optimal product feature, not a residual the intermediary fails to eliminate. Moreover, we endogenise the subscription decision through belief-confirmation preferences, creating a feedback loop between audience composition and content design that is absent in their single-journalist-to-broad-readership model. A further structural difference is that Goldman et al. (2022) assume readers take the journalist's report at face value—"credulous" traders in the spirit of Bolton et al. (2006) and Ahern and Sosyura (2014)—while our subscribers are fully Bayesian. Mispricing in our model does not rely on credulity; it arises from discrete pooling, noisy prices, endogenous participation, and the intermediary's finite reputational discipline.

Our work also connects to the broader literature on media slant and persuasion. Mullainathan and Shleifer (2005) model competing newspapers that slant toward readers' priors to attract readership; Gentzkow and Shapiro (2006) study how reputational concerns discipline or amplify media bias; and Gentzkow and Shapiro (2010) decompose media slant into demand-side (audience preferences) and supply-side (owner ideology) components. We borrow the demand-side logic—audiences value congenial content—but swap the institution: finfluencers monetise a niche rather than a mass market, and the crucial demand margin is subscription and retention rather than circulation. In the information-design tradition, Kamenica and Gentzkow (2011) formalise optimal Bayesian persuasion, showing how a sender designs an information structure to maximise the probability of a desired receiver action. Our setting adds two features absent from the canonical framework: the finfluencer jointly chooses a *signal and a fee*, creating a two-dimensional design problem; and persuasion takes place inside a *market-clearing price system* with noisy aggregation, so that the signal affects not only beliefs but also the equilibrium price that determines all agents' payoffs. The manipulation-cost parameter μ also links our model to reputational-discipline theories of media production: even when demand pushes strongly toward slant, a sender with enough credibility at stake moderates distortion. In this sense, the paper turns the traditional-media-versus-finfluencer comparison into a disciplined comparative static rather than a categorical dichotomy: legacy outlets correspond to a higher-μ region of the same model.

A third strand concerns social influence, belief-based preferences, and the formation of echo chambers. DeMarzo et al. (2003) demonstrate that agents in a network fail to account for the common origin of their neighbours' signals, producing persistent overconfidence in group views. Hirshleifer (2020) develops a theory of social transmission bias in which ideas that are simpler, more extreme, or more emotionally arousing spread faster through social networks, amplifying market-wide sentiment; Han et al. (2022) formalise how such bias affects investor behaviour. Acemoglu et al. (2010) study the spread of misinformation in social networks, showing that even rational agents can hold persistently wrong beliefs when network structure concentrates influence. We build on these insights but emphasise a distinct margin: investors experience direct *utility losses* from disconfirming information, capturing confirmation bias in the spirit of Rabin and Schrag (1999). This generates endogenous subscription selection that is absent in canonical persuasion-bias or network settings. Our approach also provides a micro-foundation for the echo-chamber phenomenon documented by Cookson et al. (2023), who find that on StockTwits self-described bulls are five times more likely to follow bullish users than bears, and that beliefs formed in these echo chambers predict lower ex-post returns and higher trading volume.

The paper also relates to the foundational literature on information aggregation and limits to arbitrage. Grossman and Stiglitz (1980) demonstrate the impossibility of fully informationally efficient markets when information acquisition is costly; Kyle (1985) shows how a single insider's information is gradually incorporated into prices through strategic trading against liquidity noise; and Shleifer and Vishny (1997) formalise limits to arbitrage when rational traders face agency frictions. De Long et al. (1990) show that noise-trader risk deters arbitrageurs, allowing mispricing to persist, while Miller (1977) demonstrates that divergence of opinion combined with short-selling constraints produces overvaluation. We introduce a new information intermediary—the finfluencer—into this classical setting. Unlike an informed trader who maximises trading profits or an analyst who provides forecasts, the finfluencer maximises *subscription revenue net of manipulation costs*. This different objective function changes the composition of traders and the informativeness of prices: standard models produce mean-unbiased prices with noise-induced variance, whereas our model produces state-dependent mispricing concentrated in the low-fundamental state because the finfluencer's pooling strategy creates an asymmetry that noisy aggregation cannot undo. Our overvaluation channel operates even without short-selling constraints, distinguishing it from the mechanism in Miller (1977).

Finally, a burgeoning empirical literature documents that social media and attention influence retail trading and asset returns. Barber and Odean (2008) and Da et al. (2011)

establish that individual attention drives retail buying behaviour; Barber et al. (2022) show that Robinhood-user attention leads to herding and temporary price increases followed by reversals; Tetlock (2007) demonstrates that media sentiment predicts short-term returns and volume; and Engelberg and Parsons (2011) provide causal evidence that local media coverage affects local trading. Peress (2014) uses newspaper strikes to identify the role of media in price formation, while Fang and Peress (2009) find that stocks with no media coverage earn higher returns. In the social-media era specifically, Cookson et al. (2024b) study attention and sentiment across Twitter, StockTwits, and Seeking Alpha, finding that sentiment predicts positive next-day returns while attention predicts negative returns; Cookson and Niessner (2020) decompose investor disagreement on social platforms into information and interpretation differences; and Pedersen (2022) models social networks as amplifiers of non-fundamental trading motives. Cookson et al. (2020) show that partisan investors' beliefs about equities diverged sharply during the COVID-19 pandemic, consistent with identity-driven processing of financial information, and Cookson et al. (2024a) document that social media can catalyse bank runs. Our paper contributes a micro-founded theoretical mechanism that generates sharp cross-sectional predictions linking confirmation-biased demand, intermediary incentives, and reputational discipline.

3 Model

We construct a four-date model of a financial market in which a revenue-maximising finfluencer intermediates between fundamentals and retail investors. The model is deliberately parsimonious: a binary fundamental, two investor types, and a single finfluencer. This tractability allows us to solve for the finfluencer's optimal slant and fee, to characterise endogenous subscription, and to derive sharp comparative statics for mispricing. We describe the environment in three parts: the timeline and fundamentals, the agents and their preferences, and the equilibrium concept. Figure 1 summarises the sequence of events, the decisions available to each agent, and the resulting payoffs.

3.1 Timeline and fundamentals

Time is indexed by $t \in \{0, 1, 2, 3\}$. At $t = 0$, the fundamental state $\theta \in \{L, H\}$ is realised with equal prior probability, $\mathbb{P}(\theta = H) = \mathbb{P}(\theta = L) = 1/2$. A single risky asset in zero net supply pays a terminal dividend d_θ at $t = 3$, where $d_H > d_L$. We write $\sigma_d \equiv \frac{1}{2}(d_H - d_L)$ for the half-spread between the high and low dividends; this quantity governs the stakes

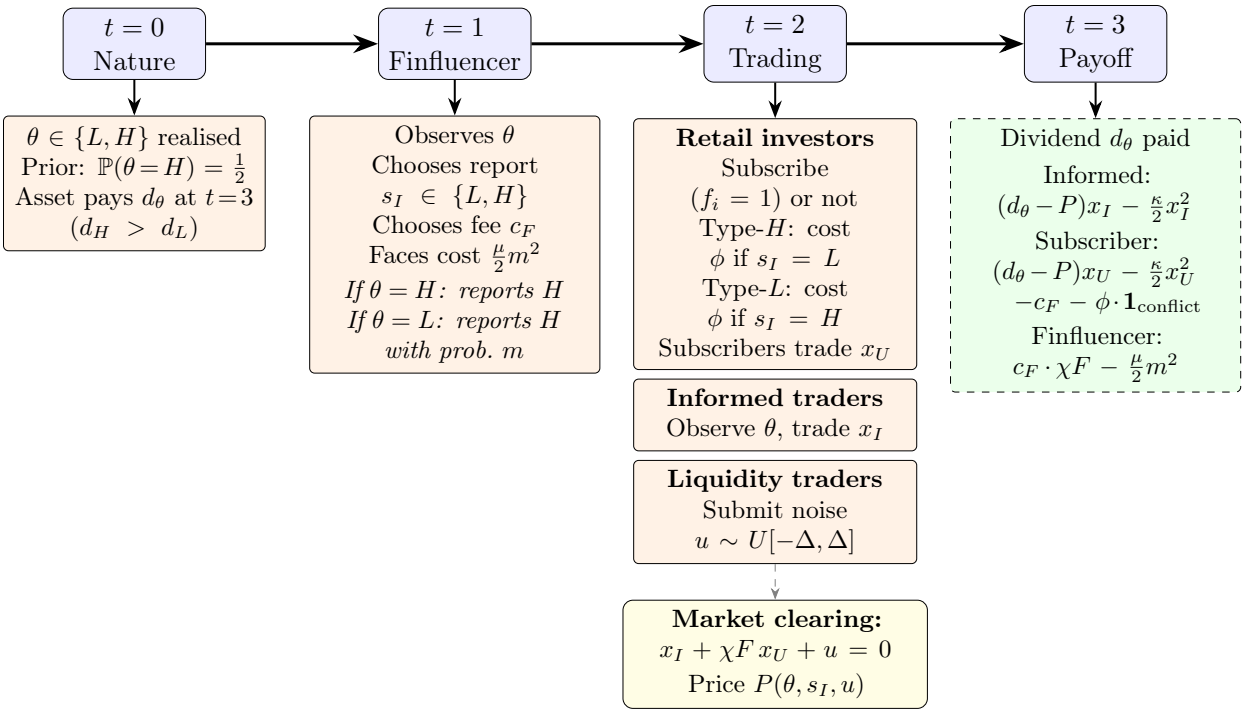


Figure 1: **Timeline, decisions, and payoffs.** At $t = 0$, nature draws the fundamental state. At $t = 1$, the finfluencer observes θ and designs both the report and the fee, while facing a convex manipulation cost. At $t = 2$, retail investors make subscription decisions and all traders submit demands; the market clears. At $t = 3$, the dividend is paid and payoffs are realised.

from mispricing. A risk-free numeraire is available in elastic supply and earns a gross return normalised to one.

At $t = 1$, the finfluencer perfectly observes θ. She then makes two choices simultaneously: a discrete report $s_I \in \{L, H\}$ and a per-subscriber monetisation rate $c_F \geq 0$. The report is communicated to all subscribers; the fee is charged upon subscription. The finfluencer's objective is to maximise total revenue net of manipulation costs,

$$R \;=\; c_F \cdot \chi F \;-\; \frac{\mu}{2} m^2, \tag{1}$$

where χ is the mass of potential retail subscribers, F is the endogenous fraction who choose to subscribe, and $\mu \geq 0$ measures reputational or implementation discipline. The quadratic form is parsimonious and captures the idea that moving from occasional slant to systematic distortion is disproportionately costly: sustaining false narratives, avoiding contradiction by future prices, and preserving credibility become harder at high m.

We allow strategic misreporting only in the bad state: if $\theta = H$ the finfluencer truthfully

reports H; if $\theta = L$ she reports H with probability $m \in [0, 1]$ and L with the complementary probability.[3] The probability m is the finfluencer's key strategic variable: it governs the extent to which she pools the bad state with the good state to cater to optimistic subscribers. When $m = 0$ reporting is fully truthful; when $m = 1$ the finfluencer always reports H regardless of the state. Note that the finfluencer observes θ at $t = 1$ and simultaneously announces both the report s_I and the fee c_F. Subscribers observe c_F before deciding whether to subscribe, but they do not observe θ directly. The finfluencer therefore chooses a state-contingent triple (θ, s_I, c_F); however, because c_F is publicly announced while θ is private information, the observable contract is the pair (s_I, c_F).

At $t = 2$, trading takes place. Retail investors decide whether to subscribe, observe the report if they do, and submit demands. Informed traders observe θ directly and trade. Liquidity traders submit exogenous demand. A market-clearing price is determined. At $t = 3$, the dividend d_θ is paid and all positions are settled.

3.2 Agents and preferences

The finfluencer. The finfluencer is a monopolistic intermediary who maximises subscription revenue net of the manipulation cost in (1). She has no trading position and no direct stake in the asset's price—her payoff comes entirely from the fees paid by subscribers, reduced by the cost of maintaining distortion. This assumption captures the institutional reality that many finfluencers earn revenue through subscriptions, paid communities, and engagement-driven advertising rather than through proprietary trading (Cookson et al., 2024c). It also provides a transparent way to discipline manipulation in the baseline model, rather than relegating credibility concerns to an auxiliary extension.

Informed traders. A unit mass of informed traders observes θ directly and trades competitively at $t = 2$. These traders are risk-neutral and face quadratic trading costs with parameter $\kappa > 0$, so that trader j's payoff from a position x_j is

$$w_j \;=\; x_j\,(d_\theta - P) \;-\; \frac{\kappa}{2}\,x_j^2.$$

[3]The restriction to one-sided misreporting is without loss of generality in our binary setting. Reporting L when $\theta = H$ would reduce the probability that the majority type-H subscribers receive a report aligned with their prior, lowering willingness to pay. The finfluencer therefore never has an incentive to "deflate" in the good state.

Optimising over x_j yields informed demand

$$x_I \;=\; \frac{d_\theta - P}{\kappa},$$

which is increasing in the gap between the dividend and the price and decreasing in the severity of trading frictions.

Uninformed retail investors. A mass $\chi > 0$ of uninformed retail investors is split into two types that capture heterogeneous priors. Type-H investors form a fraction $\lambda \in (1/2, 1)$ and hold prior $\mathbb{P}(\theta = H) = q$ with $q > 1/2$. Type-L investors form the remaining fraction $1 - \lambda$ and hold prior $\mathbb{P}(\theta = L) = q$. The assumption $\lambda > 1/2$ says that the majority of potential retail subscribers lean bullish.

Each retail investor decides whether to subscribe ($f_i = 1$) or not ($f_i = 0$). Subscription costs c_F and grants access to the finfluencer's report s_I. In addition, a subscribing investor incurs a psychological cost $\phi > 0$ whenever the report conflicts with their prior leaning: a type-H investor incurs ϕ when $s_I = L$, and a type-L investor incurs ϕ when $s_I = H$. This cost captures confirmation bias in the spirit of Rabin and Schrag (1999). When ϕ is large, investors are reluctant to subscribe to a finfluencer whose message might contradict their priors, and the finfluencer has a stronger incentive to slant toward the majority type.

Subscribers who observe s_I update their beliefs using Bayes' rule (anticipating the finfluencer's strategy m) and trade competitively, facing the same quadratic cost κ. Their demand is

$$x_U \;=\; \frac{\mathbb{E}[d_\theta \mid s_I, P] - P}{\kappa},$$

where the conditional expectation reflects both the finfluencer's report and the information in the equilibrium price.

A key assumption is that uninformed investors who do not subscribe do not participate in secondary-market trading. This is a tractable reduced form for three reinforcing frictions: fixed participation costs that deter entry without a signal, severe adverse selection against uninformed non-subscribers in a market populated by informed traders, and bounded rational attention. We relax this assumption in Section E.1 and show that the central mechanism survives.

Liquidity traders. Liquidity traders submit exogenous net demand $u \sim U[-\Delta, \Delta]$, independent of θ and of all other random variables, with $\mathbb{E}[u] = 0$ and $\mathrm{Var}(u) = \Delta^2/3$. Their role is standard:

liquidity noise prevents the price from being fully revealing, so that the finfluencer's report retains informational value even after investors observe the price.

3.3 Equilibrium concept

Definition 1 (Equilibrium)**.** An equilibrium consists of: (i) informed demand $x_I(\theta, P)$ and subscriber demand $x_U(s_I, P)$; (ii) a finfluencer strategy (m, c_F) and induced report $s_I(\theta)$; (iii) subscription decisions f_i for each retail type; and (iv) a price function $P(\theta, s_I, u)$, such that:

(i) *Optimal demands.* Given beliefs and P, demands x_I and x_U solve traders' optimisation problems.

(ii) *Optimal subscription.* Given (m, c_F) and beliefs, retail investors subscribe if and only if expected utility from subscribing (net of c_F and any confirmation cost ϕ) is nonnegative.

(iii) *Optimal content and pricing.* Given retail behaviour, the finfluencer chooses (m, c_F) to maximise revenue $R = c_F \cdot \chi F - \frac{\mu}{2} m^2$, where F is the endogenous subscriber fraction.

(iv) *Market clearing.* Total demand equals zero: $x_I + \chi F \, x_U + u = 0$.

(v) *Bayesian consistency.* Beliefs are updated using Bayes' rule wherever applicable, and all agents correctly anticipate the finfluencer's strategy m.

4 Equilibrium Pricing and Information Aggregation

Let $F \in [0, 1]$ denote the fraction of retail investors who subscribe. From market clearing and optimal demands, we obtain the equilibrium price.

Proposition 1 (Equilibrium price)**.** *For any realisation (θ, s_I, u) and subscriber fraction F,*

$$P(\theta, s_I, u) = \frac{1}{1 + \chi F} \Big(d_\theta + \kappa u + \chi F \cdot \mathbb{E}[d_\theta \mid s_I, P] \Big). \tag{2}$$

Proof. Informed traders demand $x_I = (d_\theta - P)/\kappa$. Subscribers demand $x_U = (\mathbb{E}[d_\theta \mid s_I, P] - P)/\kappa$. Substituting into $x_I + \chi F \, x_U + u = 0$ and solving for P yields (2). □

Intuitively, price is a weighted average of the informed component $(d_\theta + \kappa u)$ and subscribers' valuation $\mathbb{E}[d_\theta \mid s_I, P]$. The weight on subscribers increases with χ and F, so the finfluencer affects prices through both beliefs (via s_I) and market composition (via F).

4.1 Full revelation when $s_I = L$

Since we restrict manipulation to the bad state, observing $s_I = L$ pins down $\theta = L$ with certainty.

Corollary 1 (Price under $s_I = L$)**.** *If $s_I = L$, then $P(L, L, u) = d_L + \kappa u/(1 + \chi F)$, and $\mathbb{E}_u[P] = d_L$ (no expected mispricing).*

Proof. With $\mathbb{E}[d_\theta \mid s_I = L, P] = d_L$, equation (2) gives $P = (d_L + \kappa u + \chi F \cdot d_L)/(1 + \chi F) = d_L + \kappa u/(1 + \chi F)$. Taking expectations over u with $\mathbb{E}[u] = 0$ completes the proof. □

4.2 Pooling under $s_I = H$ and price informativeness

When $s_I = H$, subscribers face genuine uncertainty: the state may be truly high or low but misreported (with probability m). Define the price signal $s_P \equiv d_\theta + \kappa u$. Conditional on θ, s_P is uniform on $[d_\theta - \kappa\Delta,\ d_\theta + \kappa\Delta]$.

Define thresholds

$$\underline{s} \equiv d_H - \kappa\Delta, \qquad \overline{s} \equiv d_L + \kappa\Delta. \tag{3}$$

When $s_P > \overline{s}$, $\theta = H$ is revealed. When $s_P < \underline{s}$, $\theta = L$ is revealed. The overlap region $(\underline{s}, \overline{s})$ exists when $\kappa\Delta > \sigma_d$.

Lemma 1 (Price informativeness)**.** *Suppose $\kappa\Delta > \sigma_d$. The probability that s_P reveals the state is*

$$\pi_p \equiv \mathbb{P}(s_P > \overline{s} \mid \theta = H) = \mathbb{P}(s_P < \underline{s} \mid \theta = L) = \frac{\sigma_d}{\kappa\Delta} \in (0, 1). \tag{4}$$

Proof. Under $\theta = H$, $s_P \sim U[d_H - \kappa\Delta, d_H + \kappa\Delta]$. Then $\mathbb{P}(s_P > d_L + \kappa\Delta) = (d_H - d_L)/(2\kappa\Delta) = \sigma_d/(\kappa\Delta)$. The second equality follows by symmetry. □

Larger liquidity noise or costlier arbitrage reduces π_p, making prices less revealing—a standard limits-to-arbitrage force.

4.3 Subscribers' posterior under $s_I = H$

In the overlap region $s_P \in (\underline{s}, \overline{s})$, with conjecture $\hat{m}$ and equal priors, Bayes' rule gives:

$$\rho_H \equiv \mathbb{P}(\theta = H \mid s_I = H,\ s_P \in (\underline{s}, \overline{s})) = \frac{1}{1 + \hat{m}}, \tag{5}$$

since $\mathbb{P}(s_I = H \mid \theta = H) = 1$ and $\mathbb{P}(s_I = H \mid \theta = L) = \hat{m}$, while both states generate s_P values in the overlap region with equal conditional probability. Subscribers' valuation in the overlap region is thus:

$$V^*(\hat{m}) \equiv \frac{1}{1+\hat{m}}\, d_H + \frac{\hat{m}}{1+\hat{m}}\, d_L = d_L + \frac{2\sigma_d}{1+\hat{m}}. \tag{6}$$

Outside the overlap region, s_P reveals θ and the valuation equals d_θ.

Even with correct beliefs about m, the posterior in the overlap region does not collapse to 0 or 1, so the bullish report shifts valuations above d_L. Because the report is a coarse pooling rule rather than an additive distortion, there is no simple "subtract the mean bias" correction.

5 Endogenous Subscription and the Finfluencer's Problem

This section is the analytical core of the paper. We first derive the expected trading surplus that a subscriber earns under each possible report, then characterise each investor type's willingness to pay as a function of the finfluencer's manipulation probability m. We use these objects to map out the feasible fee–subscriber-mass pairs, show that type-H investors always have weakly higher willingness to pay when confirmation bias is sufficiently strong, and identify the two canonical pricing strategies open to the finfluencer.

The manipulation cost enters only the finfluencer's objective, not subscribers' utilities. This separation is useful: the demand side still determines the maximum feasible fee, while the supply side trades off catering rents against a convex cost of distortion. For fixed manipulation probability m, the fee problem is unchanged relative to the baseline demand system. The key economic trade-off appears in the choice of m: the finfluencer no longer rides a convex objective to a corner, because the quadratic cost directly disciplines high levels of manipulation.

Notation. Throughout this section we abbreviate

$$A(F) \;\equiv\; \frac{\sigma_d^2(\kappa\Delta - \sigma_d)}{\kappa^2\Delta\,(1+\chi F)^2},$$

so that the information-rent loss can be written compactly as

$$\Delta\Pi(m, F) = A(F)\frac{m}{(1+m)^2}.$$

Under Strategy 2 (optimists only) we write $A_2 \equiv A(\lambda)$ and under Strategy 1 (both types) we write $A_1 \equiv A(1)$.

5.1 Expected Trading Surplus

Under quadratic trading costs, a subscriber with posterior mean $\mathbb{E}[d_\theta \mid \Omega]$ facing price P earns instantaneous profit $(d_\theta - P)x_U - (\kappa/2)x_U^2$. Optimising over x_U and taking expectations yields expected *trading surplus*

$$\Pi(\Omega) \;=\; \frac{\mathbb{E}\big[(\mathbb{E}[d_\theta \mid \Omega] - P)^2\big]}{2\kappa}.$$

Surplus under $s_I = L$. When the finfluencer reports $s_I = L$, Corollary 1 shows that $\theta = L$ is revealed with certainty. Substituting the fully-revealing price into the surplus formula and integrating over liquidity noise:

$$\Pi_L(F) \;\equiv\; \mathbb{E}[\Pi(\Omega) \mid s_I = L] \;=\; \frac{\kappa\Delta^2}{6(1+\chi F)^2}. \tag{7}$$

Notice that $\Pi_L(F)$ is decreasing in F: the more subscribers there are, the more the market absorbs subscribers' demand, compressing the price impact and their information rent.

Surplus under $s_I = H$. When $s_I = H$, the subscriber faces genuine uncertainty. Combining Equation (6) with the price structure in Equation (2), one can show:

$$\Pi_H(m, F) \;\equiv\; \mathbb{E}[\Pi(\Omega) \mid s_I = H] \;=\; \Pi_L(F) \;-\; \Delta\Pi(m, F), \tag{8}$$

where the *information-rent loss* from pooling is

$$\Delta\Pi(m, F) \;\equiv\; \frac{m\,\sigma_d^2(\kappa\Delta - \sigma_d)}{\kappa^2\Delta\,(1+m)^2\,(1+\chi F)^2} \;=\; A(F)\frac{m}{(1+m)^2}. \tag{9}$$

Three properties of $\Delta\Pi$ are central to all subsequent results:

(i) $\Delta\Pi > 0$ *for all* $m \in (0, 1]$*:* pooling always costs subscribers some information rent.

(ii) $\Delta\Pi$ *is increasing in* m*:* more manipulation leads to a more distorted posterior and a larger forecast error.

(iii) $\Delta\Pi$ *is concave in* m*:* the marginal information-rent loss is increasing at first but eventually flattens because the report becomes almost fully pooled.

5.2 Type-Dependent Subscription Utilities and Critical Fees

A retail investor of type $k \in \{H, L\}$ subscribes if and only if their expected utility net of fees and psychological costs is nonnegative. The two types differ in their probability of receiving an aligned report.

Lemma 2 (Subscription utilities and critical fees)**.** *Let $m \in [0,1]$ and $F \in \{1-\lambda, \lambda, 1\}$ be given. Define $p_H(m) \equiv q + m - qm$ as the probability that type-H receives report H, and $p_L(m) \equiv 1 - q + qm$ as the probability that type-L receives report H. The expected utilities of subscribing are:*

$$\mathrm{EU}_H(m, c_F) \;=\; \Pi_L(F) - p_H(m)\,\Delta\Pi(m,F) - c_F - (1-q)(1-m)\,\phi, \tag{10}$$

$$\mathrm{EU}_L(m, c_F) \;=\; \Pi_L(F) - p_L(m)\,\Delta\Pi(m,F) - c_F - p_L(m)\,\phi. \tag{11}$$

Setting these equal to zero yields the critical (maximum willingness-to-pay) fees:

$$c_F^H(m) \;\equiv\; \Pi_L(F) - p_H(m)\,\Delta\Pi(m,F) - (1-q)(1-m)\,\phi, \tag{12}$$

$$c_F^L(m) \;\equiv\; \Pi_L(F) - p_L(m)\,\Delta\Pi(m,F) - p_L(m)\,\phi. \tag{13}$$

Proof. For type-H, the unconditional probability of receiving $s_I = H$ is $q \cdot 1 + (1-q)\cdot m = p_H(m)$, and the probability of a conflicting report is $(1-q)(1-m)$. Expected trading surplus is $p_H(m)\Pi_H + (1-p_H(m))\Pi_L = \Pi_L - p_H(m)\Delta\Pi$. Subtracting c_F and the expected psychological cost gives (10). The type-L expression is analogous. □

Lemma 3 (Type-H has higher willingness to pay)**.** *For any $m \in (0,1)$ and ϕ satisfying*

$$\phi \;>\; \phi^{\dagger}(m) \;\equiv\; \frac{(p_H(m) - p_L(m))\,\Delta\Pi(m,F)}{m}, \tag{14}$$

we have $c_F^H(m) > c_F^L(m)$.

Proof. Subtracting Equation (13) from Equation (12) and simplifying yields

$$c_F^H - c_F^L = \big[p_L(m) - p_H(m)\big]\Delta\Pi(m,F) + m\phi.$$

The second term is strictly positive. The condition (14) guarantees that it dominates the first term. □

5.3 Feasible Fee Schedules and Subscription Regions

Proposition 2 (Subscription regions)**.** *Under the condition of Lemma 3 ($c_F^H > c_F^L$):*

$$F(c_F, m) = \begin{cases} 1 & \text{if } c_F \le c_F^L(m), \\ \lambda & \text{if } c_F^L(m) < c_F \le c_F^H(m), \\ 0 & \text{if } c_F > c_F^H(m). \end{cases} \tag{15}$$

Proof. Type-H subscribes iff $c_F \le c_F^H$ and type-L subscribes iff $c_F \le c_F^L$. Since $c_F^H > c_F^L$, the three fee regions correspond to: both types subscribe, only type-H subscribes, and no one subscribes. □

5.4 The Finfluencer's Revenue Problem

The finfluencer maximises revenue net of manipulation costs,

$$R(c_F, m) = c_F \cdot \chi F(c_F, m) - \frac{\mu}{2} m^2.$$

Since Proposition 2 shows that F is constant within each fee interval, revenue is linear in c_F for fixed m within each interval, so the optimum within each interval is at the upper boundary fee. The manipulation cost does not depend on the fee, so the optimal *fee* conditional on m is unchanged.

Proposition 3 (Optimal fee given m: two pricing strategies)**.** *For any fixed $m \in [0, 1]$, the finfluencer's revenue-maximising fee is either:*

Strategy 1 (both types): $c_{F,1}^*(m) = c_F^L(m), \quad F_1 = 1,$

$$R_1(m) = \chi\, c_F^L(m) - \frac{\mu}{2} m^2. \tag{16}$$

Strategy 2 (type-H only): $c_{F,2}^*(m) = c_F^H(m), \quad F_2 = \lambda,$

$$R_2(m) = \chi\lambda\, c_F^H(m) - \frac{\mu}{2} m^2. \tag{17}$$

The finfluencer's optimal fee is $c_{F,k}^*(m)$ for whichever strategy $k \in \{1, 2\}$ achieves $\max\{R_1(m), R_2(m)\}$.

Proof. Revenue is strictly increasing in c_F within each participation region, while the

manipulation cost is independent of c_F. Hence the optimum within each region is at the right endpoint, exactly as in the zero-cost benchmark. □

Outside option. If both boundary revenues are negative, the finfluencer can instead set a prohibitive fee and obtain zero subscribers and zero revenue. Throughout the equilibrium analysis below, we focus on the positive-participation region in which at least one of the two boundary revenues is nonnegative, so the relevant comparison is between Strategy 1 and Strategy 2.

5.5 The Critical Threshold ϕ^*

Let $m_k^* \equiv \arg\max_{m\in[0,1]} R_k(m)$ for $k = 1, 2$ and define $R_k^* \equiv R_k(m_k^*)$.

Proposition 4 (Equilibrium strategy and threshold ϕ^*)**.** *Fix $\mu \geq 0$ and suppose the positive-participation region is such that Strategy 1 weakly dominates for sufficiently low ϕ, while Strategy 2 strictly dominates for sufficiently high ϕ. Then there exists a threshold $\phi^*(\mu) \in (0, \infty)$ such that:*

(i) If $\phi < \phi^*(\mu)$, the finfluencer uses Strategy 1.

(ii) If $\phi > \phi^*(\mu)$, the finfluencer uses Strategy 2.

(iii) If the crossing is unique, then $\phi^*(\mu)$ is weakly increasing in μ.

Proof. The strategy-specific value functions $R_1^*(\phi, \mu)$ and $R_2^*(\phi, \mu)$ are continuous in (ϕ, μ) because each is the maximum of a continuous objective over the compact interval $[0, 1]$. Under the stated crossing condition, continuity implies that the set where the two strategies deliver the same revenue is non-empty, so a threshold $\phi^*(\mu)$ exists. If the crossing is unique, then increasing μ lowers the value of Strategy 2 while leaving Strategy 1 weakly less affected, so the crossing point can only move weakly to the right. □

The threshold ϕ^* reflects a fundamental tension in the finfluencer's revenue problem. When confirmation bias is mild, pessimists tolerate occasional disconfirming messages in exchange for the signal's informational value, and the finfluencer finds it profitable to serve the full market at a moderate fee. As confirmation bias intensifies, the gap between optimists' and pessimists' willingness to pay widens until it becomes more profitable to abandon pessimists entirely, charge a steep fee to optimists, and slant aggressively toward their priors. Reputational discipline does not eliminate this logic; it moderates how far the finfluencer rides it.

Higher μ shifts the threshold ϕ^* to the right because Strategy 2 requires positive manipulation to be worthwhile—excluding pessimists only pays if the higher fee charged to optimists compensates for the lost subscriber mass, and that compensation requires slanting the report. When distortion is costly, this calculus tips less readily in favour of the optimists-only strategy. A more disciplined intermediary therefore serves a broader audience over a wider range of confirmation-bias intensities, and even when she does switch to Strategy 2, she manipulates less aggressively once there.

5.6 Optimal Manipulation Under Strategy 2

Under Strategy 2, the finfluencer chooses manipulation $m \in [0,1]$ to maximize

$$R_2(m) \;=\; \chi\lambda\, c_F^H(m) - \frac{\mu}{2}m^2. \tag{18}$$

Let $G_2(m;\phi,\mu) \equiv R_2'(m)$ denote the net marginal benefit of manipulation.

Proposition 5 (Existence and uniqueness of interior manipulation)**.** *There exists a threshold level of reputational discipline $\bar{\mu_2} > 0$ such that for all $\mu > \bar{\mu_2}$, the objective $R_2(m)$ is strictly concave on [0,1]. Under this condition:*

(i) an interior solution $m^* \in (0,1)$ exists if and only if

$$G_2(0;\phi,\mu) > 0 \qquad \textit{and} \qquad G_2(1;\phi,\mu) < 0, \tag{19}$$

(ii) in which case the optimal manipulation choice m is the unique solution to

$$G_2(m^*;\phi,\mu) = 0 \tag{20}$$

Sufficiency versus necessity. The bound $\bar{\mu_2} > 0$ is a global sufficient condition, not a necessary one. A smaller value of μ can still generate local concavity at the relevant stationary point. We impose the stronger global bound because it is easy to verify and rules out the curvature problem on the entire choice set.

Corollary 2 (Zero-discipline benchmark in the main text)**.** *Suppose Strategy 2 is considered with $\mu = 0$. Then any interior stationary point solving the rescaled cubic equation for the zero-discipline benchmark is a minimum of the Strategy 2 objective. The true optimum is bang-bang:*

$$m^* = 1 \quad \textit{if} \quad \frac{A_2}{4} < (1-q)\phi, \qquad m^* = 0 \quad \textit{if} \quad \frac{A_2}{4} \geq (1-q)\phi.$$

Proof. Immediate from Section H, which computes $R''_{2,0}(m)$ directly and shows that the Cardano root of the zero-discipline cubic is a minimum. □

The boundary conditions in (36) have a natural economic reading. At $m = 0$, the marginal gain from introducing a small amount of manipulation is positive whenever confirmation bias is strong enough relative to the information-rent loss: optimists' willingness to pay rises faster than their informational surplus falls. At $m = 1$, the reputational cost μ directly suppresses the marginal gain, ensuring that full manipulation is not optimal. Between these two boundaries, the strict concavity of R_2 guaranteed by $\mu > \bar{\mu}_2$ pins down a unique interior optimum. Economically, as manipulation intensifies the marginal reputational and implementation cost eventually dominates the incremental revenue from catering, and it is this tension that produces a stable interior manipulation choice.

It is convenient to record the boundary values explicitly:

$$G_2(0; \phi, \mu) = \chi\lambda\big[(1-q)\phi - A_2 q\big], \tag{21}$$

$$G_2(1; \phi, \mu) = \chi\lambda\left[(1-q)\phi - \frac{A_2(1-q)}{4}\right] - \mu. \tag{22}$$

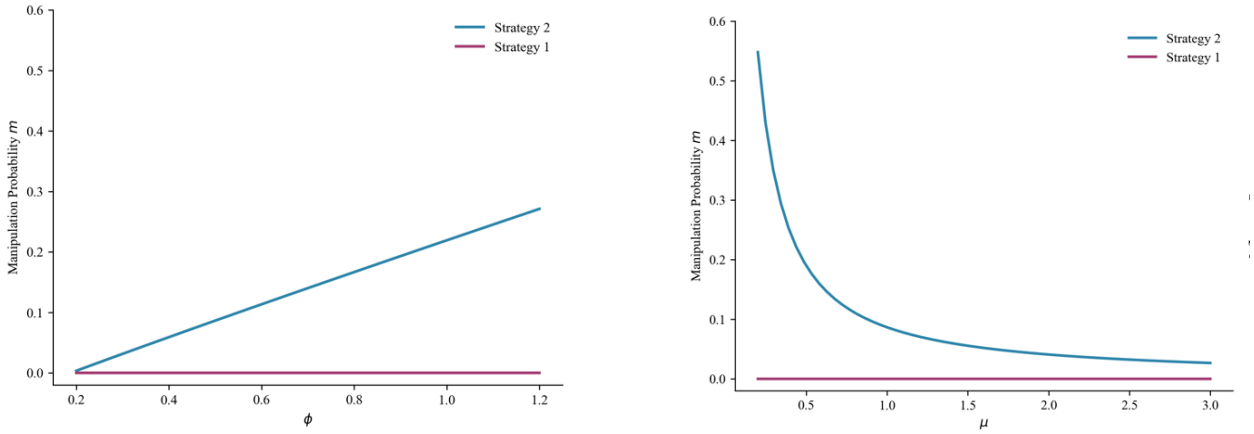
Equation (21) says that the marginal benefit of moving away from full truthfulness is larger when confirmation bias is stronger and the information-rent term A_2 is weaker. Equation (22) says that the upper boundary is disciplined directly by μ: sufficiently high reputational cost prevents the finfluencer from running to $m = 1$.

Proposition 6 (Manipulation increases in confirmation bias and decreases in discipline)**.** *Suppose Strategy 2 is optimal and the conditions of Proposition 5 hold. Then the unique interior solution m_2^* satisfies*

$$\frac{\partial m_2^*}{\partial \phi} = -\frac{\chi\lambda(1-q)}{R''_2(m_2^*)} \;>\; 0, \tag{23}$$

$$\frac{\partial m_2^*}{\partial \mu} = \frac{m_2^*}{R''_2(m_2^*)} \;<\; 0. \tag{24}$$

Proof. Apply the implicit function theorem to $G_2(m_2^*; \phi, \mu) = 0$. Since $G_{2,m} = R''_2(m_2^*) < 0$, the signs follow immediately from $G_{2,\phi} = \chi\lambda(1-q) > 0$ and $G_{2,\mu} = -m_2^* < 0$. □



(A) Effect of confirmation bias ϕ

(B) Effect of reputational discipline μ

Figure 2: Manipulation intensity. This figure plots the equilibrium manipulation intensity m^* in Proposition 6 against the psychological cost ϕ (left panel) and the reputational cost μ (right panel). Parameters: $\kappa = 2$, $\Delta = 1.2$, $\lambda = 0.7$,$\sigma_d = 1$,$d_L = 10$. Left panel: $\mu = 1$; Right panel: $\phi = 0.5$.

Figure 2 provides a visual representation of the comparative static results derived in Proposition 6. The left panel shows that the equilibrium manipulation probability m^* is strictly increasing in the confirmation bias intensity ϕ. As the psychological cost of receiving belief-inconsistent information increases, optimistic investors become more reluctant to subscribe to a finfluencer who might deliver disconfirming messages. The finfluencer responds by increasing the manipulation probability m—the likelihood of reporting H when the true state is L—to cater to the majority type's priors.

The right panel demonstrates that the manipulation probability is strictly decreasing in the reputational parameter μ. A higher value of μ implies greater reputational capital, tighter platform monitoring, or higher costs of information verification, all of which increase the marginal cost of manipulation. Consequently, finfluencers subject to stronger reputational constraints optimally choose a lower degree of manipulation.

Together, these comparative statics deliver a sharp empirical prediction: the manipulation probability varies systematically across information intermediaries, depending on their reputational exposure and the strength of confirmation bias among their audience. Finfluencers with low accountability and those catering to strongly biased investors display the highest incidence of manipulation, whereas reputable traditional media serving more diverse investor bases engage in the least informational distortion.

Proposition 6 has a direct institutional reading. Higher confirmation bias makes catering more valuable and therefore raises manipulation. Higher reputational discipline makes

aggressive slant more expensive and therefore reduces it. Low-accountability intermediaries—pseudonymous accounts, lightly moderated platforms, ephemeral chat groups—correspond to a low-μ environment; traditional media outlets, licensed analysts, and verified creators with public track records correspond to a high-μ environment. The parameter μ thus captures the key structural difference between these intermediary types within a single unified framework.

5.7 Strategy 1 and the Full Equilibrium

Under Strategy 1, the fee is set at the type-L participation constraint and $F = 1$:

$$R_1(m) = \chi c_F^L(m) - \frac{\mu}{2} m^2.$$

Because a higher m both worsens type-L's informational surplus and raises their expected psychological cost, the derivative $R_1'(m)$ is negative on $[0, 1]$; the quadratic cost only reinforces this.

Proposition 7 (No manipulation under Strategy 1)**.** *If the finfluencer serves both types, then $m_1^* = 0$.*

Proof. Differentiating R_1 yields

$$R_1'(m) = \chi \left[-p_L(m) \frac{\partial \Delta\Pi(m, 1)}{\partial m} - q\, \Delta\Pi(m, 1) - q\, \phi \right] - \mu m,$$

which is strictly negative for all $m \geq 0$. □

We can now summarise the equilibrium.

Theorem 1 (Full equilibrium characterisation)**.** *Fix $\mu \geq 0$ and suppose the single-crossing threshold $\phi^*(\mu)$ from Proposition 4 exists.*

(i) If $\phi \leq \phi^*(\mu)$, the finfluencer uses Strategy 1:

$$m^* = 0, \qquad c_F^* = c_F^L(0), \qquad F^* = 1.$$

Expected mispricing equals zero.

(ii) If $\phi > \phi^*(\mu)$ and the conditions of Proposition 5 hold, the finfluencer uses Strategy 2:

$$m^* = m_2^*, \qquad c_F^* = c_F^H(m_2^*), \qquad F^* = \lambda,$$

where m_2^* is the unique interior solution to $G_2(m;\phi,\mu)=0$.

Proof. Case (i) follows from Proposition 7. Case (ii) follows from Proposition 5 and the optimal-fee characterisation in Proposition 3. □

The equilibrium characterisation leaves pricing, posterior beliefs, subscriber masses, and the mispricing formulas analytically transparent, while treating the manipulation rate as the solution to a strictly concave one-dimensional problem. It is useful to keep two layers of selection distinct. The threshold $\phi^*(\mu)$ determines whether the finfluencer abandons pessimists and enters Strategy 2 at all. The curvature bound $\bar{\mu}_2$ together with the boundary sign conditions determines whether the Strategy 2 problem has a valid interior solution rather than a corner. Raising μ pushes the first threshold to the right while simultaneously making the within-Strategy 2 problem better behaved once that regime is entered. Taken together, the model has three economically distinct regimes: truthful reporting under low confirmation bias, bang-bang manipulation when discipline is negligible, and smooth interior manipulation once discipline is strong enough to regularise the Strategy 2 objective.

The equilibrium highlights two distinct margins that govern manipulation. The threshold $\phi^*(\mu)$ determines whether the finfluencer serves the entire market or instead targets the optimistic segment only. Conditional on entering this second regime, the manipulation choice solves a well-behaved one-dimensional optimisation problem that balances the informational rents from catering to optimistic subscribers against the increasing discipline cost of distortion. The equilibrium manipulation rate is therefore pinned down implicitly by the first-order condition of this problem.

These two margins jointly generate three economically distinct regimes. When confirmation bias is weak, the finfluencer serves both types and reports truthfully. When confirmation bias is strong enough that pessimists are excluded but discipline is negligible, manipulation occurs at the boundary. When reputational discipline is sufficiently strong, however, manipulation becomes interior and adjusts smoothly with the underlying parameters.

5.8 Comparative statics beyond ϕ and μ

The two most robust and economically salient comparative statics are those in Proposition 6. Additional primitives operate mainly through the effective information-rent term A_2. Since G_2 is decreasing in A_2, any parameter that raises A_2 reduces manipulation; any parameter that lowers A_2 increases manipulation.

Corollary 3 (Comparative statics through the information-rent term)**.** *Suppose Strategy 2 is optimal and the conditions of Proposition 5 hold.*

(i) If a parameter change raises A_2, then m_2^* falls.

(ii) If a parameter change lowers A_2, then m_2^* rises.

(iii) In particular, stronger price informativeness and larger dividend dispersion (when they raise A_2) discipline manipulation, while a noisier environment that lowers A_2 increases manipulation.

Proof. Apply the implicit function theorem to $G_2(m_2^*; \phi, \mu, A_2) = 0$. Since

$$\frac{\partial G_2}{\partial A_2} = -\chi\lambda \frac{q + m_2^*(2 - 3q)}{(1 + m_2^*)^3} < 0,$$

and $R_2''(m_2^*) < 0$, the sign of $\partial m_2^*/\partial A_2$ is negative. □

6 Mispricing Analysis

We now characterise how the finfluencer's reporting strategy distorts equilibrium prices. Define *mispricing* as the proportional deviation of the equilibrium price from the asset's true fundamental value:

$$\mathrm{MP}(\theta, s_I, u) \;\equiv\; \frac{P(\theta, s_I, u) - d_\theta}{d_\theta},$$

and let $\overline{\mathrm{MP}}(\theta, s_I) \equiv \mathbb{E}_u[\mathrm{MP}(\theta, s_I, u)]$ denote expected mispricing. Because $\mathbb{E}[u] = 0$, noise-trader demand generates no systematic bias; any non-zero $\overline{\mathrm{MP}}$ is therefore attributable entirely to the informational distortions induced by the finfluencer's report.

Three scenarios arise depending on the true state θ and the influencer's report s_I: (i) $\theta = H$, $s_I = H$; (ii) $\theta = L$, $s_I = H$; and (iii) $\theta = L$, $s_I = L$.

Proposition 8 (State-Contingent Mispricing)**.** *Under the equilibrium characterised in Proposition 1, the expected mispricing in each scenario is:*

$$\overline{\mathrm{MP}}(H, H) = -\frac{2(1 - \pi_p)\, m\, \chi F\, \sigma_d}{d_H(1 + m)(1 + \chi F)} \;\le\; 0, \tag{25}$$

$$\overline{\mathrm{MP}}(L, H) = \;\; \frac{2(1 - \pi_p)\, \chi F\, \sigma_d}{d_L(1 + m)(1 + \chi F)} \;>\; 0, \tag{26}$$

$$\overline{\mathrm{MP}}(L, L) = 0. \tag{27}$$

Proof. See Section C. □

Several features of Proposition 8 deserve emphasis.

Scenario (iii): no systematic mispricing. When the finfluencer reports $s_I = L$, she does so truthfully by construction. Subscribers immediately infer $\theta = L$, the price is centred on d_L, and $\overline{\text{MP}}(L, L) = 0$.

Scenario (i): underpricing of good news. Even when the true state is high and the finfluencer reports truthfully, the market underprices the asset on average whenever $m > 0$. Rational subscribers know that $s_I = H$ may be a manipulated signal, and therefore discount the report, pushing the equilibrium price below d_H.

Scenario (ii): overpricing under manipulation. The central case is $\theta = L$, $s_I = H$. The manipulated high report inflates subscribers' posterior above d_L, and this distorted demand is transmitted into prices. Overpricing is increasing in χF and decreasing in m on a per-event basis because higher manipulation is more aggressively discounted. But when equilibrium manipulation rises, the pooling event also happens more often. This distinction between per-event severity and event frequency is what matters for aggregate expected mispricing.

6.1 Aggregate Expected Mispricing

Aggregating across scenarios yields the following ex-ante result.

Proposition 9 (Aggregate Mispricing)**.** *The ex-ante expected mispricing satisfies:*

$$\overline{\text{MP}} \;=\; \frac{(1-\pi_p)\, m\, \chi F\, \sigma_d}{(1+m)(1+\chi F)} \left(\frac{1}{d_L} - \frac{1}{d_H} \right) \;>\; 0 \qquad \textit{for all } m > 0. \tag{28}$$

Proof. Direct computation, weighting Equations (25)–(27) by their respective ex-ante probabilities. □

Proposition 9 establishes that even partially strategic reporting generates a strictly positive aggregate price distortion. Intuitively, the overpricing in low-fundamental states more than offsets the underpricing in high-fundamental states in proportional terms because the same dollar distortion is amplified when scaled by the smaller fundamental d_L.

Corollary 4 (Comparative Statics of Aggregate Mispricing)**.** *Conditional on Strategy 2:*

(i) $\overline{\text{MP}}$ *is strictly increasing in* χF*.*

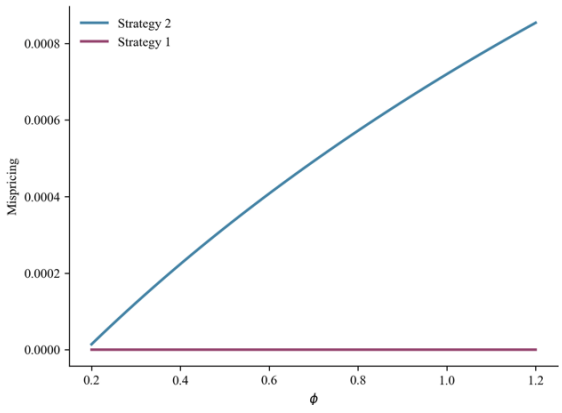
(ii) $\overline{\text{MP}}$ *is strictly increasing and concave in* m *through the factor* $m/(1+m)$*.*

(iii) Since $\partial m_2^*/\partial\phi > 0$ *and* $\partial m_2^*/\partial\mu < 0$*, aggregate expected mispricing is increasing in* ϕ *and decreasing in* μ*.*

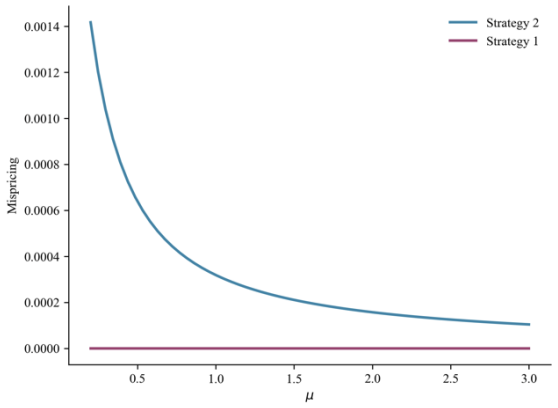
(iv) $\overline{\text{MP}}$ *is strictly decreasing in* π_p*.*

Proof. The first, second, and fourth statements follow directly from Equation (28). The third combines (28) with Proposition 6. □

The parameter μ enters prices only through equilibrium manipulation: it disciplines the finfluencer's incentive problem without altering the market-clearing or Bayesian-updating blocks of the model. The pricing and mispricing formulas therefore remain analytically transparent, and the model delivers a clean comparative statement: platforms, business models, and institutional settings with higher reputational discipline generate less manipulation and therefore less mispricing.



(A) Effect of confirmation bias ϕ on mispricing



(B) Effect of reputational discipline μ on mispricing

Figure 3: Aggregate expected mispricing. This figure plots the aggregate expected mispricing $\overline{\text{MP}}$ against the psychological cost ϕ (left panel) and the reputational cost μ (right panel). Parameters: $\kappa = 2$, $\Delta = 1.2$, $\lambda = 0.7$,$\sigma_d = 1$,$d_L = 10$. Left panel: $\mu = 1$; Right panel: $\phi = 0.5$.

Figure 3 illustrates how the aggregate expected mispricing $\overline{\text{MP}}$ varies with the two key structural parameters under Strategy 2, integrating the results from Corollary 4 and the equilibrium manipulation responses established in Proposition 6.

The left panel of Figure 3 shows that aggregate expected mispricing $\overline{\text{MP}}$ is strictly increasing in the confirmation bias intensity ϕ. As established in Corollary 4(iii), higher ϕ increases equilibrium manipulation m^* (Proposition 6), which in turn elevates mispricing through the factor $m^*/(1+m^*)$ in Equation (28). Intuitively, stronger confirmation bias intensifies the finfluencer's catering incentive, leading to more aggressive pooling of bad states with bullish reports and consequently greater overvaluation in low-fundamental states.

The right panel demonstrates that aggregate expected mispricing is strictly decreasing in the reputational parameter μ. A higher μ reduces equilibrium manipulation m^*, which directly mitigates mispricing through the same transmission channel. Economically, stronger reputational discipline—stemming from platform monitoring, formal verification requirements, or long-term reputation concerns—deters the finfluencer from distorting information, thereby reducing the resulting price inefficiency. The downward-sloping pattern indicates that institutional environments with greater accountability promote more informative asset prices.

7 Welfare and Policy

Even with Bayesian traders, finfluencer monetisation can reduce welfare. We measure welfare as expected terminal consumption net of real trading distortions and the psychological costs borne by subscribers. Because we interpret μ as an expected continuation-value loss or credibility cost borne by the intermediary outside the current trading round, we treat it as a private disciplining device rather than a contemporaneous resource cost in the welfare accounting below. It still matters for welfare through its effect on equilibrium $m^\star$. Appendix N shows how the accounting changes if one instead treats μ as a direct resource cost and discusses the regulator's trade-off in choosing discipline intensity.

7.1 Welfare decomposition

Aggregate welfare decomposes as

$$W = W_{\text{alloc}} - W_{\text{trade}} - W_{\text{psych}}. \tag{29}$$

Subscription fees are transfers between the finfluencer and her subscribers and cancel from aggregate welfare. The welfare-relevant distortions are therefore mispricing—which misallocates capital by inducing trades at prices that deviate from fundamentals—and the psychological costs borne by subscribers who receive a report conflicting with their private

signal.

Before deriving the welfare-loss terms, we clarify the treatment of informed trader surplus. Informed traders observe θ directly; their demand $x_I = (d_\theta - P)/\kappa$ and resulting gross surplus $(d_\theta - P)^2/(2\kappa)$ are affected by price distortions induced by the finfluencer. However, because the risky asset is in *zero net supply*, every unit purchased by a buyer has a corresponding seller: gains accruing to informed buyers come at the expense of informed or subscribing sellers, and vice versa. Consequently, the *sum* of informed trader surpluses integrates to zero across the full population in expectation. Under manipulation, the price distortion redistributes surplus between informed traders and subscribers, but this redistribution is a *transfer* and does not represent a deadweight loss. Subscription fees are likewise transfers between the finfluencer and her subscribers and therefore cancel from aggregate welfare. The welfare-relevant distortions are the real resource costs: quadratic trading costs incurred on mispriced positions, and psychological disconfirmation costs borne by subscribers.

Proposition 10 (Welfare effects of finfluencer manipulation)**.** *The welfare loss relative to truthful reporting is*

$$\Delta W \;=\; \underbrace{\frac{m^\star}{2}\cdot(1-\pi_p)\cdot\frac{\chi F}{1+\chi F}\cdot\frac{2\sigma_d^2}{(1+m^\star)^2\kappa}}_{\textit{mispricing-induced trading cost}} \;+\; \underbrace{\chi F\cdot\bar{\phi}(m^\star)}_{\textit{expected psychological cost}}\quad, \tag{30}$$

where

$$\bar{\phi}(m^\star) \;=\; \tfrac{1}{2}(1-q)\phi \;+\; \tfrac{1}{2}\big[m^\star q + (1-m^\star)(1-q)\big]\phi \tag{31}$$

is the average expected psychological cost per subscriber.

Proof. See Section D. □

The welfare decomposition in equation (30) reveals a structural complementarity between the two distortion channels: stronger confirmation bias raises $m^\star$ through the finfluencer's first-order condition and inflates the direct psychological cost per subscriber. A larger subscriber base amplifies both the mass of distorted trading and the aggregate psychological burden. Lower price informativeness expands the overlap region where subscribers cannot learn the true state from prices, compounding the trading-cost distortion.

The comparative static in μ has a direct policy interpretation. Because μ reduces $m^\star$, stronger reputational discipline attenuates both the mispricing term and the psychological-cost term in (30). In other words, the same force that disciplines manipulation also improves equilibrium welfare.

7.2 Policy implications

The welfare decomposition maps directly onto the design of regulatory interventions.

Monetisation transparency and track records. Requiring finfluencers to disclose subscription revenue, affiliate commissions, and sponsorship relationships raises subscriber scepticism about the influencer's reporting incentives. Mandatory track-record reporting and performance dashboards raise the effective manipulation cost μ by making distortion more reputationally expensive. In our model, both effects discipline manipulation.

Improving price informativeness. Reducing barriers to informed trading, relaxing short-selling constraints, and expanding market access for arbitrageurs each raise π_p, directly shrinking the overlap region where mispricing survives. This intervention attacks the first term in (30) multiplicatively via $(1-\pi_p)$: even if manipulation persists, higher price informativeness prevents subscribers from being systematically misled.

A caution on access-expansion policies. Policies that merely expand access to finfluencer content—for instance, subsidising financial literacy through social media channels or incentivising content creation in underserved communities—increase χF without correcting the influencer's incentives. Both terms in (30) are increasing in χF, so such policies amplify welfare losses rather than reducing them.

8 Conclusion and Policy Implications

This paper develops a tractable equilibrium theory of financial influencer behaviour and its consequences for asset pricing and welfare. We close a gap in the literature by providing a formal framework to answer a question that existing theory was not equipped to address: *can social media influencers generate genuine first-moment mispricing—overvaluation in expectation—even when their audience invests rationally?*

Our answer is yes, and the mechanism is precise. In the interior-manipulation region of the parameter space, the finfluencer's slant takes the form of a discrete probability $m^\star \in (0,1)$—the probability of reporting H when the true state is L—rather than a continuous additive slant. Because Bayesian updating is non-linear in this probability, investors cannot fully de-bias the signal even when they correctly conjecture $m^\star$ in equilibrium. Combined with liquidity noise that prevents prices from fully revealing the state, this generates persistent expected overvaluation in the pooling region. By contrast, the zero-discipline benchmark produces bang-bang manipulation rather than a stable interior choice. That benchmark is useful for

anonymous fringe environments, but the paper's baseline focuses on the empirically richer interior region created by positive reputational discipline.

A central feature of the model is the convex manipulation cost, interpreted as reputational discipline, platform scrutiny, or the effort required to sustain deception. This term delivers a well-behaved finfluencer objective, yields a unique interior manipulation choice under transparent parameter restrictions, and maps directly to institutions: low-accountability finfluencers manipulate more, while high-reputation traditional media manipulate less.

The paper also delivers several additional results. First, the equilibrium characterisation provides a transparent solution for the manipulation probability, subscription fee, and equilibrium price function as explicit or implicitly defined functions of the underlying structural parameters—confirmation bias intensity ϕ, reputational discipline μ, signal precision q, market depth κ, and subscriber mass χ. Second, the mispricing formulas yield a set of testable cross-sectional predictions: assets followed by lower-discipline, higher-reach intermediaries should exhibit larger return reversals and greater overvaluation in bad states. A particularly sharp prediction is an accountability gradient across platforms and creator types: pseudonymous or weakly monitored venues should display more aggressive bullish pooling and stronger reversals than high-accountability venues with track records, verification, and institutional oversight. Third, the welfare decomposition quantifies the social cost of the prevailing influencer business model—subscription-funded, audience-catering content—and identifies the parameters along which interventions are most effective.

Policy implications and future research. The welfare analysis points directly to the design of regulatory interventions. Three classes of policy emerge from the model's comparative statics. Monetisation transparency and verified track records work by raising the effective cost of distortion. Improving price informativeness through more aggressive arbitrage and better disclosure reduces the region in which manipulation can survive Bayesian updating. By contrast, policies that expand access to commercially incentivised influencer content without altering incentives can worsen welfare.

This paper opens several productive avenues for future research. On the theoretical side, the model abstracts from influencer competition; a market-structure extension with multiple finfluencers catering to overlapping subscriber pools could examine whether competition disciplines manipulation or instead intensifies the race to cater. A dynamic extension with endogenous reputation accumulation—where past accuracy is publicly observable and feeds back into μ—would characterise the long-run equilibrium between disclosure and credibility. On the empirical side, the model's predictions are sharp and testable using influencer-level

panel data matched to returns and retail order flow.

References

Acemoglu, D., Ozdaglar, A., and ParandehGheibi, A. (2010). Spread of (mis)information in social networks. *Games and Economic Behavior*, 70(2):194–227.

Ahern, K. R. and Sosyura, D. (2014). Who writes the news? Corporate press releases during merger negotiations. *Journal of Finance*, 69(1):241–291.

Allcott, H., Braun, S., and Gentzkow, M. (2022). Digital addiction. *American Economic Review*, 112(7):2424–2463.

Barber, B. M., Huang, X., Odean, T., and Schwarz, C. (2022). Attention-induced trading and returns: Evidence from Robinhood users. *Journal of Finance*, 77(6):3141–3190.

Barber, B. M. and Odean, T. (2008). 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.

Bolton, P., Scheinkman, J. A., and Xiong, W. (2006). Executive compensation and short-termist behaviour in speculative markets. *Review of Economic Studies*, 73(3):577–610.

Cookson, J. A., Engelberg, J., and Mullins, W. (2023). Echo chambers. *Review of Financial Studies*, 36(2):450–500.

Cookson, J. A., Engelberg, J. E., and Mullins, W. (2020). Does partisanship shape investor beliefs? Evidence from the COVID-19 pandemic. *Review of Asset Pricing Studies*, 10(4):863–893.

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., and Schiller, C. (2024a). Social media as a bank run catalyst. *Journal of Financial Economics*. Forthcoming.

Cookson, J. A., Lu, R., Mullins, W., and Niessner, M. (2024b). The social signal. *Journal of Financial Economics*, 158:103870.

Cookson, J. A., Mullins, W., and Niessner, M. (2024c). Social media and finance. *Oxford Research Encyclopedia of Economics and Finance*.

Cookson, J. A. and Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *Journal of Finance*, 75(1):173–228.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5):1461–1499.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, 66(1):67–97.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance*, 64(5):2023–2052.

Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2):280–316.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Goldman, E., Martel, J., and Schneemeier, J. (2022). A theory of financial media. *Journal of Financial Economics*, 145(1):239–258.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408.

Han, B., Hirshleifer, D., and Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, 57(1):390–412.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22(3):477–498.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *Journal of Finance*, 75(4):1779–1831.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *Journal of Finance*, 32(4):1151–1168.

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053.

Pedersen, L. H. (2022). Game on: Social networks and markets. *Journal of Financial Economics*, 146(3):1097–1119.

Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *Journal of Finance*, 69(5):2007–2043.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics*, 114(1):37–82.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*, 52(1):35–55.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock

market. *Journal of Finance*, 62(3):1139–1168.

A Proofs and Derivations

A.1 Derivation of the equilibrium price (Proposition 1)

Informed traders have $\Omega_I = \{\theta\}$, so $\mathbb{E}[d_\theta \mid \Omega_I] = d_\theta$ and $x_I = (d_\theta - P)/\kappa$. Subscribers have $\Omega_U = \{s_I, P\}$ and demand $x_U = (\mathbb{E}[d_\theta \mid s_I, P] - P)/\kappa$. Market clearing requires

$$\frac{d_\theta - P}{\kappa} + \chi F \cdot \frac{\mathbb{E}[d_\theta \mid s_I, P] - P}{\kappa} + u = 0.$$

Multiplying through by κ and solving for P yields (2).

A.2 Expected trading surplus and the information-rent loss

When $s_I = L$, the state is fully revealed and the subscriber's demand is

$$x_U = \frac{d_L - P}{\kappa} = -\frac{u}{1 + \chi F}.$$

Substituting into quadratic-trading surplus and integrating over $u \sim U[-\Delta, \Delta]$ gives

$$\Pi_L(F) = \frac{\kappa \Delta^2}{6(1 + \chi F)^2}.$$

When $s_I = H$, subscribers are uncertain in the overlap region and use posterior mean $V^*(m) = d_L + \frac{2\sigma_d}{1+m}$. The same algebra as in the main text gives

$$\Pi_H(m, F) = \Pi_L(F) - A(F)\frac{m}{(1+m)^2},$$

which implies (9). Differentiating,

$$\frac{\partial \Delta\Pi}{\partial m} = A(F)\frac{1-m}{(1+m)^3}, \qquad \frac{\partial^2 \Delta\Pi}{\partial m^2} = -2A(F)\frac{2-m}{(1+m)^4} < 0.$$

Thus the information-rent loss is increasing and concave in m.

B Additional Proofs for Section 5

B.1 Proof of Proposition 4

By Proposition 7, Strategy 1 is optimised at $m = 0$, so its value is

$$R_1^*(\phi, \mu) = \chi\big[\Pi_L(1) - (1-q)\phi\big],$$

which is continuous, strictly decreasing in ϕ, and independent of μ. Strategy 2 also defines a continuous value function because its objective is continuous on the compact set $[0, 1]$. Moreover, for any (ϕ, μ) the finfluencer can always choose $m = 1$ under Strategy 2, so

$$R_2^*(\phi, \mu) \geq \chi\lambda \left[\Pi_L(\lambda) - \frac{A_2}{4}\right] - \frac{\mu}{2}.$$

Under the stated condition in Corollary 2, $\Pi_L(\lambda) - A_2/4 > \mu/2$, $R_2^*(\phi, \mu) \geq R_2(1) > 0$ for all ϕ.

Since $R_1^*(\phi, \mu) \to -\infty$ as $\phi \to \infty$ while $R_2^*(\phi, \mu)$ is bounded below by the positive constant $R_2(1)$, there exists a finite $\tilde{\phi}$ such that $R_2^*(\phi, \mu) > R_1^*(\phi, \mu)$ for all $\phi > \tilde{\phi}$. For sufficiently small ϕ, Strategy 1 weakly dominates because the type-L participation constraint is easily satisfied by a moderate fee while manipulation yields negligible catering gains. By continuity of both value functions on the compact set $[0, \bar{\phi}_1]$, the two functions must cross, yielding the threshold $\phi^*(\mu)$.

If the crossing is unique, which holds when $R_2^* - R_1^*$ is monotone in φ, a condition verified numerically in Appendix M for the baseline calibration, then increasing μ shifts R_2^* weakly downward (it lowers the value of manipulation) while leaving R_1^* unchanged, so $\phi^*(\mu)$ moves weakly to the right. □

B.2 Proof of Proposition 5

Under Strategy 2, $F = \lambda$ and the finfluencer maximises $R_2(m) = \chi\lambda\, c_F^H(m) - \frac{\mu}{2}m^2$. Using Equations (12) and (9), we can write

$$c_F^H(m) \;=\; \Pi_L(\lambda) - A_2 \frac{[q + m(1-q)]\, m}{(1+m)^2} - (1-q)(1-m)\,\phi. \tag{32}$$

Differentiating Equation (32) and substituting into $R_2'(m)$ gives

$$G_2(m;\phi,\mu) \;\equiv\; \frac{\partial R_2}{\partial m} = \chi\lambda\left[(1-q)\phi - A_2\frac{q+m(2-3q)}{(1+m)^3}\right] - \mu\, m. \tag{33}$$

An interior optimum solves $G_2(m;\phi,\mu) = 0$.

Suppose Strategy 2 is optimal. Then:

(i) The second derivative of R_2 is

$$R_2''(m) = \chi\lambda A_2\frac{2[(3q-1)+m(2-3q)]}{(1+m)^4} - \mu. \tag{34}$$

(ii) If

$$\mu \;>\; \bar{\mu}_2 \;\equiv\; 2\chi\lambda A_2(3q-1), \tag{35}$$

then R_2 is strictly concave on $[0,1]$.

(iii) Under strict concavity, an interior solution $m_2^* \in (0,1)$ exists if and only if

$$G_2(0;\phi,\mu) > 0 \qquad \text{and} \qquad G_2(1;\phi,\mu) < 0, \tag{36}$$

in which case the solution is unique.

Part (i) follows by direct differentiation. Therefore its maximum is attained at $m = 0$ and equals $2(3q-1)$, which gives the sufficient concavity bound (35). Once R_2 is strictly concave, R_2' is strictly decreasing and hence has a unique zero in $(0,1)$ iff $R_2'(0) > 0 > R_2'(1)$.

For part (ii), note that the term multiplying $\chi\lambda A_2$ in (34) is decreasing in m for $q > 1/2$. The term in (34) multiplying $\chi\lambda A_2$ is decreasing in m:

$$\frac{d}{dm}\left(\frac{2[(3q-1)+m(2-3q)]}{(1+m)^4}\right) = \frac{6[2-5q-m(2-3q)]}{(1+m)^5} < 0$$

for all $q > 1/2$ and $m \in [0,1]$ so its maximum over $[0,1]$ is attained at $m = 0$ and equals $2(3q-1)$. Therefore (35) is sufficient for $R_2''(m) < 0$ on the whole interval. Under strict concavity, the first derivative G_2 is strictly decreasing. Hence (36) is necessary and sufficient for a unique zero in $(0,1)$.

B.3 Proof of Proposition 6

Let $G_2(m;\phi,\mu)=0$ define the interior optimum. Then

$$\frac{\partial m^*}{\partial \phi}=-\frac{G_{2,\phi}}{G_{2,m}}=-\frac{\chi\lambda(1-q)}{R_2''(m^*)}>0,$$

and

$$\frac{\partial m^*}{\partial \mu}=-\frac{G_{2,\mu}}{G_{2,m}}=-\frac{-m^*}{R_2''(m^*)}=\frac{m^*}{R_2''(m^*)}<0,$$

because $R_2''(m^*)<0$.

B.4 Proof of Proposition 7

Under Strategy 1,

$$R_1(m)=\chi\left[\Pi_L(1)-A_1\frac{[1-q+qm]m}{(1+m)^2}-(1-q+qm)\phi\right]-\frac{\mu}{2}m^2.$$

Differentiating gives

$$R_1'(m)=\chi\left[-A_1\frac{(1-q)+m(3q-1)}{(1+m)^3}-q\phi\right]-\mu m<0,$$

because every term on the right-hand side is non-positive and the first two are strictly negative for $q>1/2$. Hence $m_1^*=0$.

C Proof of Proposition 8

When $\theta=H$ and $s_I=H$, there are two relevant regions. In the revealing region, subscribers infer $\theta=H$ and price equals $d_H+\kappa u/(1+\chi F)$, which has zero conditional mean mispricing. In the overlap region, subscribers use posterior valuation $\bar{d}(m)=\frac{d_H+md_L}{1+m}$, so

$$P(H,H,u)=\frac{d_H+\kappa u+\chi F\bar{d}(m)}{1+\chi F}.$$

Subtracting d_H and taking expectations over the overlap region yields

$$\overline{\mathrm{MP}}(H,H)=-\frac{2(1-\pi_p)m\chi F\sigma_d}{d_H(1+m)(1+\chi F)}.$$

When $\theta = L$ and $s_I = H$, the same logic yields

$$P(L, H, u) = \frac{d_L + \kappa u + \chi F \bar{d}(m)}{1 + \chi F}$$

in the overlap region, and hence

$$\overline{\mathrm{MP}}(L, H) = \frac{2(1 - \pi_p)\chi F \sigma_d}{d_L(1 + m)(1 + \chi F)}.$$

Finally, when $s_I = L$, the state is revealed and expected mispricing is zero.

D Proof of Proposition 10

Conditional on the pooling event ($\theta = L, s_I = H$), subscribers' posterior valuation is $V^* = d_L + \frac{2\sigma_d}{1+m^\star}$ and the equilibrium price is

$$P^{m^\star}(u) = \frac{d_L + \kappa u + \chi F V^*}{1 + \chi F}.$$

Relative to truthful reporting, the valuation wedge is

$$\delta \equiv V^* - d_L = \frac{2\sigma_d}{1 + m^\star}.$$

The excess quadratic trading cost generated by this wedge is

$$\frac{\chi F\, \delta^2}{2\kappa(1 + \chi F)}.$$

Substituting $\delta = 2\sigma_d/(1 + m^\star)$ gives

$$\frac{2\chi F \sigma_d^2}{\kappa(1 + \chi F)(1 + m^\star)^2}.$$

Weighting by the probability of the pooling event, $\frac{m^\star}{2}$, and by the probability that the price is not revealing, $(1 - \pi_p)$, yields the first term of (30). The expected psychological cost per subscriber follows by averaging conflict probabilities across the high and low states, yielding (31).

E Extensions

E.1 Allowing non-subscribers to trade

Suppose non-subscribers trade with attenuated intensity

$$x_{NS} = \gamma \frac{\mathbb{E}_0[d_\theta] - P}{\kappa}, \qquad \gamma \in [0, 1].$$

Market clearing becomes

$$x_I + \chi F x_U + \chi(1 - F) x_{NS} + u = 0,$$

so the price formula generalises to

$$P = \frac{d_\theta + \kappa u + \chi F \mathbb{E}[d_\theta | s_I, P] + \chi(1 - F)\gamma \mathbb{E}_0[d_\theta]}{1 + \chi F + \chi(1 - F)\gamma}.$$

The mechanism survives. Non-subscriber demand dilutes the impact of subscribers' distorted beliefs, but does not eliminate it unless γ is very close to one and the subscriber mass is negligible. Hence the main results continue to hold with attenuated magnitudes.

E.2 Endogenous reputation accumulation

The static parameter μ can be micro-founded by repeated interaction. Suppose that after the current period, the finfluencer loses continuation value V_{future} with probability $\rho(m)$, where $\rho'(m) > 0$ and $\rho''(m) \geq 0$. A second-order Taylor approximation around low manipulation levels gives an expected continuation loss proportional to m^2, which maps directly into the reduced-form term $\frac{\mu}{2} m^2$. This interpretation clarifies why traditional media, regulated brokers, and audited research shops plausibly have high μ: a false report damages brand capital built over many periods.

E.3 Competition among multiple finfluencers

With multiple finfluencers, two forces work in opposite directions. A disciplining force arises when subscribers can compare historical accuracy across creators; in that case, the effective manipulation cost μ rises because the reputational consequences of slant become more severe. A differentiation force arises when creators specialise in increasingly extreme niches; then

effective confirmation bias rises because each creator serves a more homogeneous audience. Which force dominates is an empirical question and likely platform specific.

E.4 Platform algorithm as a primitive

Model algorithmic amplification as a multiplier $\gamma_A \geq 1$ on the effective subscriber mass, so the price-impact term becomes $\chi\gamma_A F$. Higher γ_A does not change the subscriber's Bayesian problem, but magnifies the transmission of distorted beliefs into prices. The interaction between γ_A and μ is especially relevant empirically: a platform that both amplifies engagement-heavy content and imposes little reputational discipline is exactly the environment in which our model predicts the largest mispricing.

F Continuous Additive Slant Benchmark

To clarify why our discrete-pooling mechanism generates mispricing while a naive additive-bias model does not, consider a benchmark in which the finfluencer reports

$$s_I = d_\theta + b$$

for a deterministic bias b known in equilibrium. Rational subscribers subtract the bias:

$$\mathbb{E}[d_\theta \mid s_I] = s_I - b = d_\theta.$$

The report therefore contains no residual distortion once the bias is anticipated, and prices equal fundamentals in expectation. In our model, by contrast, the report is not an additive perturbation of a continuous signal. It is a coarse, state-contingent pooling rule with endogenous probability m. Bayesian updating on such a rule cannot be undone by subtracting a constant. That is why expected mispricing survives even with fully rational investors.

G Detailed derivation of the non-revealing-region surplus

This appendix spells out the algebra behind the information-rent loss term in Equation (9). The logic is useful because it clarifies exactly where the loss comes from and why it is naturally concave in the manipulation probability.

G.1 Price regions under a bullish report

Conditional on $s_I = H$, there are three regions of the price signal $s_P = d_\theta + \kappa u$.

1. **High revealing region:** $s_P > \overline{s} = d_L + \kappa\Delta$. This can only happen when $\theta = H$, so subscribers infer the state correctly.
2. **Low revealing region:** $s_P < \underline{s} = d_H - \kappa\Delta$. This can only happen when $\theta = L$, so subscribers again infer the state correctly.
3. **Overlap region:** $s_P \in (\underline{s}, \overline{s})$. Both states are possible, and subscribers use the posterior $\rho_H = 1/(1+m)$.

Only the third region creates a wedge between subscribers' posterior valuation and fundamentals. In the revealing regions, the report and price jointly identify the state, so the finfluencer's slant has no first-order effect on posterior beliefs.

G.2 Posterior valuation and forecast errors

In the overlap region, subscribers value the asset at

$$V^*(m) = \frac{d_H + m d_L}{1+m} = d_L + \frac{2\sigma_d}{1+m}.$$

The valuation wedge relative to each state is therefore

$$\eta_H(m) \equiv V^*(m) - d_H = -\frac{2m\sigma_d}{1+m}, \tag{37}$$

$$\eta_L(m) \equiv V^*(m) - d_L = \frac{2\sigma_d}{1+m}. \tag{38}$$

The wedge is negative in the high state and positive in the low state: a bullish report is discounted when the state is truly high but overvalued when the state is truly low. This asymmetry is the seed of the mispricing results in the main text.

Conditional on (θ, s_I, u) in the overlap region, the equilibrium price is

$$P(\theta, H, u) = \frac{d_\theta + \kappa u + \chi F V^*(m)}{1 + \chi F}. \tag{39}$$

Hence the subscriber's demand in the overlap region is

$$x_U^{\text{overlap}} = \frac{V^*(m) - P}{\kappa} = \frac{\eta_\theta(m) - \kappa u}{\kappa(1+\chi F)}. \tag{40}$$

G.3 Expected surplus in the overlap region

Substituting (40) into the quadratic-trading payoff,

$$(d_\theta - P)x_U - \frac{\kappa}{2}x_U^2,$$

and simplifying yields

$$\Pi^{\text{overlap}}(\theta, m, F, u) = \frac{(\eta_\theta(m) - \kappa u)^2}{2\kappa(1+\chi F)^2}. \tag{41}$$

Taking expectations over the overlap region therefore requires three ingredients:

$$\mathbb{E}[\eta_\theta(m)^2 \mid s_I = H,\ \text{overlap}], \qquad \mathbb{E}[u^2 \mid \text{overlap}], \qquad \mathbb{E}[\eta_\theta(m)u \mid s_I = H,\ \text{overlap}].$$

The first term is easy. Under a bullish report in the overlap region, the posterior probabilities of the states are $1/(1+m)$ and $m/(1+m)$, so

$$\begin{aligned}\mathbb{E}[\eta_\theta(m)^2 \mid s_I = H,\ \text{overlap}] &= \frac{1}{1+m}\eta_H(m)^2 + \frac{m}{1+m}\eta_L(m)^2 \\ &= \frac{1}{1+m}\frac{4m^2\sigma_d^2}{(1+m)^2} + \frac{m}{1+m}\frac{4\sigma_d^2}{(1+m)^2} \\ &= \frac{4m\sigma_d^2}{(1+m)^2}.\end{aligned} \tag{42}$$

The second term is the variance of the liquidity shock over the relevant truncated interval. The third term is more subtle. In the low state, the overlap region is centred symmetrically around zero and so the conditional mean of u is zero. In the high state, however, the overlap region is skewed because the left tail is longer than the right tail. As a result, the covariance term $\mathbb{E}[\eta_\theta(m)u \mid s_I = H,\ \text{overlap}]$ is not identically zero. This is exactly why a naive shortcut that ignores truncation understates the information-rent loss.

G.4 Recovering the reduced-form expression

Carrying out the truncated-uniform integrations and assembling the pieces from (41) produces the reduced-form expression

$$\Pi_H(m, F) = \Pi_L(F) - A(F)\frac{m}{(1+m)^2},$$

where

$$A(F) = \frac{\sigma_d^2(\kappa\Delta - \sigma_d)}{\kappa^2\Delta(1+\chi F)^2}.$$

The factor $\kappa\Delta - \sigma_d$ is the width of the overlap region measured in price-signal units. The term $1/(1+\chi F)^2$ appears because a larger subscriber base pushes more of the posterior distortion into price rather than into private informational rents.

Two useful implications follow immediately. First, $A(F)$ is smaller when the subscriber base is larger, because subscribers compete away part of the informational rent they collectively create. Second, for fixed F, the function $m/(1+m)^2$ is hump shaped and concave on $[0,1]$, so the information-rent loss rises with manipulation but at a diminishing rate. This concavity is what makes the supply-side quadratic cost especially natural: one curvature comes from informational degradation, the other from reputational discipline.

G.5 Why the main model only needs the reduced form

The main text uses the compact formula (9) rather than the raw truncated-integral expression for two reasons. First, all subscription decisions depend on the report only through the pair $(\Pi_L, \Delta\Pi)$. Second, once the finfluencer's optimisation problem is written in terms of $A(F)$, the model cleanly separates demand-side primitives (confirmation bias and participation) from the supply-side disciplining force μ. The reduced form is therefore not merely a simplification: it makes the mechanism transparent.

H Zero-discipline benchmark and the corner-solution problem

This appendix characterises the zero-discipline benchmark in which the manipulation-cost term is shut down. The exercise clarifies why that benchmark fails to produce a valid interior optimum under Strategy 2 and why positive reputational discipline yields a better-behaved manipulation choice.

H.1 The objective when $\mu = 0$

Set $\mu = 0$ and suppose the finfluencer targets only type-H investors. Her objective is then

$$R_{2,0}(m) = \chi\lambda\, c_F^H(m),$$

with

$$c_F^H(m) = \Pi_L(\lambda) - A_2 \frac{[q + m(1-q)]m}{(1+m)^2} - (1-q)(1-m)\phi.$$

Differentiating once gives

$$R'_{2,0}(m) = \chi\lambda \left[(1-q)\phi - A_2 \frac{q + m(2-3q)}{(1+m)^3} \right].$$

This is the first-order condition for the zero-discipline benchmark.

H.2 Rescaling and the zero-discipline cubic

Multiplying the bracketed first-order condition by $(1+m)^3/A_2$ gives

$$q + m(2-3q) - B_{2,0}(1+m)^3 = 0, \qquad B_{2,0} \equiv \frac{(1-q)\phi}{A_2}.$$

This is the cubic equation governing the zero-discipline benchmark. Writing down its Cardano root does not resolve the economic issue, because the stationary point lies on a convex objective.

Differentiating again yields

$$R''_{2,0}(m) = \chi\lambda A_2 \frac{2[(3q-1) + m(2-3q)]}{(1+m)^4}. \tag{43}$$

Because $q > 1/2$, the numerator in (43) is strictly positive for all $m \in [0,1]$. Therefore $R''_{2,0}(m) > 0$ on the entire interval. The supposedly "interior optimum" from the zero-cost first-order condition is thus a *minimum*, not a maximum.

H.3 The bang-bang benchmark

Once the interior candidate is recognised as a minimum, the true solution to the zero-discipline problem must lie at a boundary:

$$m^\star_{2,0} \in \{0, 1\}.$$

The two corners are easy to compute:

$$R_{2,0}(0) = \chi\lambda \left[\Pi_L(\lambda) - (1-q)\phi \right], \tag{44}$$

$$R_{2,0}(1) = \chi\lambda \left[\Pi_L(\lambda) - \frac{A_2}{4} \right]. \tag{45}$$

Hence

$$m^{\star}_{2,0} = \begin{cases} 1, & \text{if } \dfrac{A_2}{4} < (1-q)\phi, \\ 0, & \text{if } \dfrac{A_2}{4} \geq (1-q)\phi. \end{cases} \tag{46}$$

The economic content of (46) is immediate. Without reputational discipline, the finfluencer either never manipulates or always manipulates. There is no reason to stop in the interior. That is why the zero-cost model is ill suited for a baseline in which manipulation intensity is meant to be an endogenous and economically interpretable choice.

H.4 Why quadratic discipline is a natural baseline

The baseline model includes the term $\frac{\mu}{2}m^2$ in the finfluencer's objective. This does three things simultaneously.

1. It preserves the demand system: subscriber willingness to pay, posterior beliefs, price aggregation, and the mispricing formulas all remain unchanged.
2. It restores a genuine interior optimum by directly offsetting the positive curvature in (43).
3. It delivers a clean economic comparative static: higher reputational discipline lowers manipulation.

In other words, the specification is targeted: it changes the supply-side problem exactly where curvature is needed, without rewriting the rest of the market structure.

H.5 The baseline model nests the benchmark

It is useful to view the baseline model as nesting the zero-discipline benchmark. As $\mu \downarrow 0$, the interior optimum drifts toward the bang-bang solution in (46). This interpretation matters conceptually because the zero-discipline case is a knife-edge environment in which the finfluencer faces no escalating discipline for sustained distortion. Once even a small amount of convex discipline is introduced, the economically relevant interior solution reappears.

H.6 A useful robustness interpretation

Equation (46) can still be useful as a robustness benchmark. It says that in extremely low-accountability environments, one should expect a much more discontinuous manipulation

pattern. That benchmark is empirically plausible for anonymous pump-and-dump channels, ephemeral chat groups, or platforms with almost no moderation. The baseline model therefore produces a natural spectrum:

$$\text{high } \mu \implies \text{low, interior manipulation,} \qquad \text{low } \mu \implies \text{bang-bang manipulation.}$$

This spectrum is arguably more realistic than a model in which every intermediary, regardless of institutional setting, has the same interior manipulation formula.

I Primitive comparative statics and local sign maps

The main text emphasises the clean comparative statics with respect to ϕ and μ. This appendix shows how the remaining primitives enter the interior solution through the first-order condition

$$G_2(m;\phi,\mu,A_2) = \chi\lambda\left[(1-q)\phi - A_2\frac{q+m(2-3q)}{(1+m)^3}\right] - \mu m = 0.$$

Under strict concavity, the sign of $\partial m^\star/\partial\xi$ is the sign of $-G_{2,\xi}/R_2''(m^\star)$.

I.1 The information-rent term as a sufficient statistic

A convenient way to organise the comparative statics is through

$$A_2 = \frac{\sigma_d^2(\kappa\Delta - \sigma_d)}{\kappa^2\Delta(1+\chi\lambda)^2}.$$

Any parameter that increases A_2 raises the informational value of truthful reporting and therefore disciplines manipulation. Any parameter that decreases A_2 weakens the information-rent loss from slant and therefore encourages manipulation.

This representation is especially useful because it keeps the supply-side term μm separate from the demand-side object A_2. The model thus has two intuitive levers:

1. **Catering benefits**, captured by $(1-q)\phi$ and the size of the optimist audience.
2. **Distortion costs**, captured jointly by A_2 and μ.

I.2 Liquidity noise Δ

Holding other parameters fixed,

$$A_2 = \frac{\sigma_d^2}{(1+\chi\lambda)^2}\left(\frac{1}{\kappa} - \frac{\sigma_d}{\kappa^2\Delta}\right).$$

Hence

$$\frac{\partial A_2}{\partial \Delta} = \frac{\sigma_d^3}{\kappa^2\Delta^2(1+\chi\lambda)^2} > 0. \tag{47}$$

More liquidity noise makes truthful reports more valuable because the market price becomes less able to correct distortions on its own. Since $\partial m^\star/\partial A_2 < 0$, Equation (47) implies

$$\frac{\partial m^\star}{\partial \Delta} < 0.$$

This sign is the opposite of a purely "noise helps manipulators" intuition and is worth stressing: in our setting, more noise increases the private value of information to subscribers, so manipulation becomes more costly in willingness-to-pay terms.

I.3 Trading cost κ

Differentiating A_2 with respect to κ gives

$$\frac{\partial A_2}{\partial \kappa} = \frac{\sigma_d^2(2\sigma_d - \kappa\Delta)}{\kappa^3\Delta(1+\chi\lambda)^2}. \tag{48}$$

Therefore:

$$\frac{\partial m^\star}{\partial \kappa}\begin{cases} < 0, & \text{if } \kappa\Delta < 2\sigma_d, \\ > 0, & \text{if } \kappa\Delta > 2\sigma_d. \end{cases}$$

The effect of κ is genuinely conditional. When trading costs are low enough, raising κ mainly reduces the profitability of informed trading and therefore weakens the price discipline that a truthful report generates. But when trading costs are already high, a further increase in κ also compresses the private value of the subscriber's information, which can lower the marginal gain from manipulation. The net effect is therefore conditional.

I.4 Dividend dispersion σ_d

Differentiating A_2 with respect to σ_d yields

$$\frac{\partial A_2}{\partial \sigma_d} = \frac{\sigma_d(2\kappa\Delta - 3\sigma_d)}{\kappa^2\Delta(1+\chi\lambda)^2}. \tag{49}$$

Hence A_2 rises with σ_d whenever $\sigma_d < \frac{2}{3}\kappa\Delta$, which is the empirically relevant region when the price is imperfectly revealing but the overlap region is not tiny. In that case,

$$\frac{\partial m^\star}{\partial \sigma_d} < 0.$$

Larger dividend dispersion makes truthful information more valuable and thereby disciplines slant.

I.5 Audience scale $\chi\lambda$

The audience variables are more nuanced because they affect both the scale of revenues and the dilution of informational rents. Let $x \equiv \chi\lambda$. Then

$$G_2(m) = x(1-q)\phi - \frac{xC}{(1+x)^2}\frac{q + m(2-3q)}{(1+m)^3} - \mu m,$$

where $C = \sigma_d^2(\kappa\Delta - \sigma_d)/(\kappa^2\Delta)$. Differentiating with respect to x gives

$$\frac{\partial G_2}{\partial x} = (1-q)\phi - C\frac{q + m(2-3q)}{(1+m)^3}\frac{1-x}{(1+x)^3}. \tag{50}$$

Equation (50) shows that the effect of audience scale is generally positive once $x \geq 1$, because the second term then becomes weakly positive and the first term is strictly positive. For small audiences the sign can be ambiguous. Economically, two forces operate in opposite directions:

1. A larger optimist audience increases the revenue payoff from catering.
2. A larger subscriber mass pushes more of the distorted posterior into prices and away from private information rents.

In most applications of interest—large retail communities on social platforms—the first force should dominate.

I.6 Belief precision q

The effect of q is also ambiguous. Differentiating G_2 with respect to q at fixed m yields

$$\frac{\partial G_2}{\partial q} = -\chi\lambda\phi - \chi\lambda A_2 \frac{1-3m}{(1+m)^3}. \tag{51}$$

The first term is negative: when investors are more convinced of their priors, the mass of conflict-sensitive subscribers shrinks. The second term can be positive or negative depending on whether m is below or above $1/3$. This is a formal way of stating a simple intuition. More dogmatic priors can either intensify catering incentives or make truthful information more valuable, depending on where the equilibrium already sits.

I.7 What remains unambiguously signed

Not all primitive comparative statics are unambiguous. The two clearest signs are

$$\frac{\partial m^\star}{\partial \phi} > 0, \qquad \frac{\partial m^\star}{\partial \mu} < 0.$$

These are precisely the two that are most central for the paper's narrative and most directly tied to institutional interpretation.

J Traditional media versus finfluencers

The parameter μ (reputational cost) makes it possible to compare different information intermediaries within a unified framework rather than treating "traditional media" and "finfluencers" as categorically different objects.

J.1 A simple ordering result

Consider two intermediaries a and b that face the same audience composition, the same market environment, and the same confirmation-bias parameter ϕ, but differ in reputational discipline:

$$\mu_a > \mu_b.$$

Suppose both choose Strategy 2 and both satisfy the conditions for an interior optimum. Then Proposition 6 implies immediately that

$$m_a^\star < m_b^\star.$$

Substituting into Equation (28) shows that the lower-discipline intermediary generates larger aggregate mispricing.

This ordering result is economically attractive because it is robust to many institutional details. The model does not require the traditional media outlet and the finfluencer to have different audiences, different signal technologies, or different market access. It is enough that one has more credibility at stake.

J.2 Interpretation of μ

Several real-world forces can be mapped into a high or low value of μ.

1. **Brand capital.** Established outlets lose future subscription and advertising revenue when accuracy suffers.
2. **Verification and legal risk.** Regulated research shops, licensed brokers, and institutional publishers face a higher probability of complaint, retraction, or sanction.
3. **Archival visibility.** Publicly searchable track records raise the cost of contradiction by future prices.
4. **Platform moderation.** Stronger moderation, fact checking, or anti-manipulation enforcement increases the expected cost of slant.

Finfluencers, by contrast, often operate under weak versions of all four constraints. Accounts can be pseudonymous, content can be rapidly deleted or buried, and engagement can be rewarded even after accuracy deteriorates. That bundle of features is exactly what a low value of μ is meant to capture.

J.3 A regime table

Table 1 makes explicit the comparative-static spectrum implicit in the model. The same basic environment can generate very different equilibrium outcomes depending on how accountable the intermediary is.

Table 1: Media-discipline regimes

Regime	Interpretation	Equilibrium manipulation	Price implication
High μ	Traditional media, licensed analysts, highly visible creators	Low or zero interior $m^\star$; truthful reporting more common	Little or no systematic overvaluation
Intermediate μ	Mainstream finfluencers, partially monitored platforms	Positive interior $m^\star$ under Strategy 2	State-dependent overvaluation and later reversal
Low μ	Anonymous channels, weak moderation, ephemeral communities	Interior solution approaches bang-bang benchmark; manipulation can be extreme	Large temporary overvaluation, especially after bullish pooling

J.4 Why the comparison matters empirically

The ordering in $m^\star$ is only half the story. A lower-discipline intermediary also affects prices more when its audience is larger or more coordinated. Hence the empirical prediction is not simply "finfluencers are worse than traditional media." It is more precise:

> Mispricing should be most severe when low-μ intermediaries reach large, homogeneous, confirmation-seeking audiences in assets with weak arbitrage discipline.

This is exactly the type of cross-sectional heterogeneity researchers can test.

K Empirical predictions and measurement

This appendix sketches how the model maps to empirical work. The goal is not to provide a full reduced-form or structural design, but to clarify which primitives correspond to observable proxies.

K.1 Measuring the model primitives

Confirmation bias ϕ. The most direct proxies are follower-level measures of ideological or sentiment homogeneity, echo-chamber structure, or self-declared bullishness on social platforms (Cookson et al., 2023; Cookson and Niessner, 2020). A creator whose audience is unusually one sided should have a higher effective ϕ in the model.

Reputational discipline μ. Potential proxies include account verification, institutional affiliation, historical permanence of content, editorial oversight, track-record disclosure, and the existence of ex post performance dashboards. Accounts connected to broker-dealers, publishers, or registered entities should map to higher μ than pseudonymous creators monetising through referrals and private chat groups.

Prediction 1 (platform-accountability gradient). Holding fixed the asset environment and the size of the reachable audience, lower-μ creators should exhibit more one-sided bullish pooling in bad states, stronger contemporaneous price pressure, and larger subsequent reversals than higher-μ creators. This is the cleanest empirical implication of the model.

Prediction 2 (discipline $\times$ reach). The effect of low μ should be strongest when effective audience size χF is large. A fringe creator with little accountability but no reach need not move prices much; a low-μ creator with a large, homogeneous audience should.

Prediction 3 (competition $\times$ discipline). Greater creator density should reduce manipulation on verification-heavy platforms, where competition effectively raises the cost of falsehood by making creators easier to compare. The same increase in creator density can raise manipulation on recommendation-heavy platforms that sort users into narrower echo chambers and thereby lower the effective discipline facing each creator.

Subscriber mass χF. Follower counts, paid-subscriber counts, repost networks, and platform-specific engagement metrics all map naturally to the effective audience size in the model. In empirical work, one would ideally measure not just raw followers but the subset that actually trades.

Price informativeness π_p. Liquidity, analyst coverage, institutional ownership, short-interest availability, option-market depth, and the speed with which earnings news moves

prices all serve as natural proxies. The model predicts that manipulation is more damaging when this discipline is weak.

K.2 Creator-level predictions

A clean creator-level prediction is:

Holding audience composition fixed, creators with lower reputational discipline should choose more manipulation and generate larger post-message reversals.

This can be taken to data by comparing verified institutional commentators, independent newsletter writers, and platform-native finfluencers that differ sharply in track-record visibility and enforcement exposure.

A second creator-level prediction is that the same account should manipulate less after events that raise its effective μ: joining a regulated platform, becoming verified, entering a sponsorship relationship that requires disclosure, or being required to publish a historical scorecard.

K.3 Asset-level predictions

The model also produces sharp cross-sectional predictions across assets.

1. Assets with high retail ownership and low shorting capacity should exhibit larger overvaluation following bullish content from low-μ creators.
2. The effect should be stronger when the creator's audience is especially homogeneous or sentiment aligned.
3. Reversals should be concentrated in the low-fundamental state, because that is where the model's pooling distortion operates.
4. The pricing effect should be strongest when external price discovery is weak, for instance in small-cap, low-coverage, or hard-to-short names.

The last point is important. The model is not simply a theory of "hype." It is a theory of hype interacting with market structure. The same post can have small effects in a highly liquid, information-rich asset and much larger effects in a thinly traded name.

K.4 Event-study design implications

Suppose one constructs a panel of influencer posts matched to stock returns. The model suggests splitting events along three dimensions:

1. **Creator discipline:** high-μ versus low-μ accounts.
2. **Audience confirmation demand:** homogeneous bullish followers versus mixed audiences.
3. **Asset arbitrage capacity:** easy-to-arbitrage versus hard-to-arbitrage stocks.

The strongest return continuation and subsequent reversal pattern should appear in the "low-μ, high-ϕ, low-π_p" cell. This three-way interaction is more distinctive than any one-dimensional prediction and provides a way to separate our mechanism from generic attention shocks.

K.5 A measurement challenge and a way forward

The practical challenge is that $m^\star$ is not directly observed. But the model suggests several ways to back it out:

1. from the historical frequency with which bullish reports are followed by later negative fundamentals;
2. from textual features that identify one-sided positive framing in situations with subsequently weak outcomes;
3. from a structural inversion of the fee, audience composition, and reversal pattern implied by the model.

The last approach is especially promising because the model makes μ and ϕ enter in distinct places: ϕ operates through subscriber willingness to pay and μ through the finfluencer's supply-side first-order condition.

K.6 Why the reputational-discipline margin helps empirics

Introducing an explicit reputational-discipline margin creates a directly observable empirical dimension. It helps distinguish why two creators facing similar audiences might nonetheless

manipulate different amounts: differences in verification, institutional affiliation, archival visibility, or enforcement exposure map into differences in μ. That makes the theory more falsifiable, not less.

L Additional extension details

This appendix expands on the qualitative extensions in Section E.

L.1 Dynamic enforcement and endogenous μ

Suppose a platform uses a strike system under which the expected penalty after manipulation level m is

$$\mathcal{P}(m) = \sum_{j=1}^{J} \pi_j(m)\ell_j,$$

where $\pi_j(m)$ is the probability of sanction type j and ℓ_j is the corresponding loss in future revenue. If the sanction probabilities are smooth and convex near the equilibrium, then a local quadratic approximation

$$\mathcal{P}(m) \approx \mathcal{P}(0) + \mathcal{P}'(0)m + \frac{\mathcal{P}''(\tilde{m})}{2}m^2$$

justifies the reduced-form term $\frac{\mu}{2}m^2$. In this interpretation, stricter platform rules increase μ by steepening the penalty schedule.

L.2 Competition and audience sorting

With multiple creators, one can distinguish between *verification competition* and *sorting competition*. Verification competition raises the effective cost of falsehood because creators can be cross-checked against one another. Sorting competition lowers the cost because each creator can specialise in a narrower echo chamber. A useful empirical implication is that creator density should reduce manipulation on search-heavy platforms but increase it on recommendation-heavy platforms.

L.3 Heterogeneous audiences within a single creator

The baseline model uses two investor types for tractability, but the logic extends to richer audience structures. If follower preferences are distributed continuously over a sentiment index, the creator's fee problem becomes a standard screening problem and the manipulation choice tilts the distribution of marginal subscribers. The quadratic cost remains valuable in that environment because it continues to guarantee a well-behaved interior manipulation decision.

L.4 Regulatory interventions

The model highlights three regulatory margins:

1. **Raise μ:** strengthen disclosure, verification, and track-record requirements.
2. **Lower ϕ:** improve media literacy and encourage cross-cutting information exposure.
3. **Raise π_p:** support arbitrage and price discovery in assets disproportionately affected by influencer traffic.

These levers are complements. Raising μ without improving price informativeness still leaves some room for distortions. Improving price informativeness without raising μ still leaves the creator with incentives to cater. The model's value is to show how the levers interact.

M Numerical illustration and calibration intuition

This appendix provides a simple numerical illustration of the mechanism. The point is not to estimate the model, but to show how reputational discipline μ changes equilibrium outcomes in a transparent way.

M.1 A baseline calibration

Take a baseline environment with

$$q = 0.65, \qquad \lambda = 0.7, \qquad \chi = 1.2, \qquad \kappa = 2, \qquad \Delta = 1.2, \qquad \sigma_d = 1.$$

These values satisfy the imperfect-revelation condition $\kappa\Delta > \sigma_d$. They also imply a nontrivial overlap region in which the bullish report cannot be fully debiased by the price.

For these values,

$$A_2 = \frac{\sigma_d^2(\kappa\Delta - \sigma_d)}{\kappa^2\Delta(1+\chi\lambda)^2}$$

is strictly positive but economically modest, reflecting that truthful subscriber information has value, yet that value is attenuated by a sizeable coordinated retail audience. Numerically, $A_2 \approx 0.086$ and the global concavity threshold from Proposition 5 is

$$\bar{\mu}_2 = 2\chi\lambda A_2(3q-1) \approx 0.137.$$

Thus a moderate manipulation-cost parameter such as $\mu = 0.20$ is enough to satisfy the transparent sufficient condition for strict concavity. The zero-discipline benchmark looks very different. If $\phi = 0.40$, then $A_2/4 \approx 0.022 < (1-q)\phi = 0.14$, so Appendix H implies the optimum is the corner $m^\star = 1$. With the same ϕ and $\mu = 0.20$, the unique interior solution to the first-order condition is approximately $m^\star = 0.52$. This illustrates how positive discipline converts the bang-bang benchmark into a smooth interior choice. Table 2 makes the calibration point more concrete.

Table 2: Illustrative manipulation choices around the baseline calibration

μ	Relative to $\bar{\mu}_2$	Global concavity?	$m^\star$ at $\phi = 0.40$	Interpretation
0	zero discipline	No	1.00	bang-bang benchmark
0.10	below $\bar{\mu}_2$	No	1.00	corner solution persists
0.14	approximately $\bar{\mu}_2$	approximately yes	0.79	interior choice emerges
0.20	above $\bar{\mu}_2$	Yes	0.52	moderate discipline regularises the problem
0.50	well above $\bar{\mu}_2$	Yes	0.18	strong discipline sharply compresses slant

The table underscores the main quantitative message: the sufficient bound is not large. In this calibration, moving from $\mu = 0$ to a moderate value around $\mu = 0.20$ is already enough to replace full manipulation with an interior level slightly above one half.

M.2 Comparative-static scenarios

The following table summarises the qualitative comparative-static patterns one should expect around the baseline.

Table 3: Illustrative parameter changes around the baseline

Parameter change	Effect on $m^\star$	Effect on mispricing	Intuition
ϕ increases	Up	Up	Catering value rises because conflict avoidance becomes more valuable
μ increases	Down	Down	Distortion is more reputationally expensive
Δ increases	Down	Ambiguous ex ante but often down	Truthful information becomes more valuable when prices are noisier
$\chi\lambda$ increases	Usually up	Up strongly	Larger coordinated audience raises the payoff from catering and increases price impact
σ_d increases	Usually down	Mixed	Truthful signals matter more when fundamentals are more dispersed

The key distinction is between *the equilibrium manipulation choice* and *the price impact of a given manipulation choice*. Even when a parameter reduces $m^\star$, it may still amplify price distortions conditional on manipulation by increasing the weight of subscribers in the market-clearing equation. This is why empirical work should separate the creator-side choice margin from the market-impact margin.

M.3 How discipline shifts the feasible interior region

Within Strategy 2, the boundary conditions $G_2(0) > 0$ and $G_2(1) < 0$ imply a lower and upper bound on ϕ. In the baseline with $\mu = 0.20$, these conditions translate approximately

into $0.160 < \phi < 0.702$. Raising μ pushes the upper bound upward because it makes the upper corner $m = 1$ less attractive, thereby helping replace the corner with an interior choice. Across strategies, however, Proposition 4 implies that higher μ also shifts the overall regime threshold $\phi^*(\mu)$ to the right. The two effects are distinct: stronger discipline makes Strategy 2 internally better behaved once entered, but it also makes the optimists-only strategy harder to justify in the first place.

M.4 An illustrative regime-threshold shift

The baseline calibration above happens to lie in a region where Strategy 2 dominates for a wide range of ϕ. To illustrate Proposition 4 more directly, consider instead a crossing case with

$$q = 0.65, \qquad \lambda = 0.60, \qquad \chi = 0.50, \qquad \kappa = 2, \qquad \Delta = 1.2, \qquad \sigma_d = 1.$$

Holding these primitives fixed, a numerical comparison of the strategy-specific value functions gives approximately

$$\phi^*(0) \approx 0.197, \qquad \phi^*(0.30) \approx 0.307.$$

Thus raising discipline from $\mu = 0$ to $\mu = 0.30$ shifts the Strategy 1/Strategy 2 boundary to the right by about 56 percent. Economically, stronger reputational discipline means that the finfluencer requires a much more confirmation-seeking audience before it becomes worthwhile to abandon pessimists and sell only to optimists.

M.5 What the calibration is meant to show

The model is not fragile. For a wide range of moderate parameter values, the conditions for strict concavity are easy to satisfy because the required lower bound on μ is finite and transparent. In other words, one does not need an implausibly large reputational penalty to obtain a valid interior solution. At the same time, once μ is small, the solution moves quickly toward the corner benchmark described in Section H. This feature is desirable: it means the model can accommodate both mainstream creators with some accountability and fringe accounts with almost none.

N Alternative welfare accounting when μ is a real resource cost

The main text interprets μ as a continuation-value loss borne by the intermediary outside the current trading round and therefore excludes it from contemporaneous welfare accounting. A reader may instead prefer to treat the manipulation cost as a direct resource cost—for example, because sustaining deception requires paid promotion, legal avoidance, or organisational effort. In that case, the welfare decomposition becomes

$$\Delta W^{\text{alt}} = \Delta W + \frac{\mu}{2}(m^\star)^2. \tag{52}$$

This alternative formulation does not change the comparative-static logic with respect to ϕ. Stronger confirmation bias still raises manipulation and therefore increases both the market-distortion term and the direct cost term. The effect of μ is subtler. A higher μ now has two opposing effects:

1. it directly raises the resource cost attached to any given level of manipulation;
2. it indirectly lowers equilibrium manipulation through Proposition 6.

In sufficiently low-manipulation regions, the second effect dominates and total welfare improves as μ rises. In very high-manipulation regions, the sign can in principle be ambiguous. This ambiguity is one reason the main text adopts the continuation-value interpretation: it keeps the policy discussion focused on the disciplining effect of reputation rather than on whether the penalty itself is socially costly.

A regulator who directly chooses discipline intensity would therefore face a genuine trade-off problem,

$$\min_{\mu \geq 0} \Delta W^{\text{alt}}(\mu),$$

where ΔW^{alt} depends on μ both directly and through the induced equilibrium manipulation level $m^\star(\mu)$. If $m^\star(\mu)$ is differentiable, then

$$\frac{d\Delta W^{\text{alt}}}{d\mu} = \frac{1}{2}\big(m^\star(\mu)\big)^2 + \left[\frac{\partial \Delta W}{\partial m^\star} + \mu m^\star(\mu)\right]\frac{dm^\star}{d\mu}.$$

Since $\frac{dm^\star}{d\mu} < 0$, the second term is the indirect welfare gain from discipline: it scales the marginal welfare damage of manipulation by the amount by which stronger discipline

compresses equilibrium slant. An interior policy optimum μ^{reg} therefore solves

$$\frac{1}{2}\big(m^\star(\mu^{\text{reg}})\big)^2 = -\left[\frac{\partial \Delta W}{\partial m^\star} + \mu^{\text{reg}} m^\star(\mu^{\text{reg}})\right] \frac{dm^\star}{d\mu}\Big|_{\mu=\mu^{\text{reg}}}.$$

This expression makes the regulator's trade-off explicit. The socially optimal level of discipline need not be infinite: once manipulation is already low, the remaining marginal gain from further discipline can be outweighed by the direct resource cost of sustaining it.

N.1 Why the distinction is empirically useful

The two interpretations can be distinguished empirically. If tighter platform rules or disclosure mandates reduce manipulation without creating obvious deadweight costs, then the continuation-value interpretation is more natural. If reducing manipulation requires substantial screening, compliance, or litigation effort, then the resource-cost interpretation may be preferable. The model is flexible enough to accommodate either reading; the core mispricing mechanism is unchanged.

O Fee comparative statics under Strategy 2

Because the manipulation cost enters the creator's problem but not subscribers' utilities, it affects equilibrium fees only indirectly through $m^\star$. Under Strategy 2,

$$c_F^H(m) = \Pi_L(\lambda) - A_2 \frac{[q + m(1-q)]m}{(1+m)^2} - (1-q)(1-m)\phi.$$

Differentiating gives

$$\frac{dc_F^H}{dm} = (1-q)\phi - A_2 \frac{q + m(2-3q)}{(1+m)^3}.$$

At the interior optimum,

$$\chi\lambda \frac{dc_F^H}{dm} = \mu m^\star.$$

Therefore the fee is increasing in $m^\star$ at the optimum: the creator manipulates up to the point where the induced increase in willingness to pay by optimistic subscribers is exactly balanced by the marginal cost of distortion. Combining this relation with Proposition 6 yields two useful corollaries:

1. a rise in μ lowers both manipulation and the equilibrium optimist-targeted fee;

2. a rise in ϕ raises both manipulation and the equilibrium fee whenever Strategy 2 remains optimal.

This gives the model a joint prediction for prices and creator monetisation: lower-discipline creators should not only manipulate more, but should also be able to charge more to a homogeneous bullish audience.