Robin Dub

How Negative Is Too Negative? Modeling the Trust Dynamics of Deinfluencing

Abstract: Negative recommendations can increase perceived influencer credibility by signaling independence from promotional incentives, yet that benefit need not increase indefinitely with the frequency of criticism. This article develops a sociological and formal account of deinfluencing in which audiences infer latent communicator types—promotional, independently evaluative, or contrarian—from the observed rate of negative judgments. The resulting credibility function is single-peaked: an independence benefit with diminishing returns is offset by an accelerating reputational cost of habitual or strategic negativity.

Keywords: deinfluencing; influencer marketing; credibility; trust; signaling; Bayesian inference; confirmation bias; strategic contrarianism

1. Introduction

Influencer credibility has traditionally been treated as a function of perceived expertise, authenticity, relatability, and trustworthiness, yet the conditions under which these perceptions are formed are considered through relatively static comparisons rather than as the outcome of repeated evaluative behaviour over time. Deinfluencing has emerged as a useful problem for understanding how credibility is produced in influencer marketing. Recent research suggests that negative recommendations can increase perceived trustworthiness by reducing skepticism about an influencer's motives. An influencer who discourages consumption demonstrates a willingness to withhold endorsement, which gives audiences evidence that commercial incentives do not fully determine their judgments. In this application, the informative value of negativity becomes the conveyed signal of independence or willingness to go against the expected action.

This article approaches that problem through two related questions. The first concerns the direction and social meaning of the effect itself: under what conditions does negative evaluation increase perceived independence, honesty, or credibility, and through what sociological processes does this occur? The second concerns the graphable form of that effect: if the frequency of negative judgments is treated as a continuous variable (as opposed to a simple contrast), what functional relationship should we expect between negativity and credibility, and why should that relationship be nonlinear rather than monotonic? These questions are necessarily connected, because the relationship cannot be justified independently of the social interpretations that generate it, while the sociological account remains incomplete unless it specifies how those interpretations are likely to change as the frequency of negative behaviour changes.

The starting assumption is that negative evaluation can function as a signal of independence because it is, under ordinary circumstances, less compatible with the image of an indiscriminately promotional communicator. An influencer who recommends every product encountered offers the audience limited grounds on which to distinguish genuine enthusiasm from commercial

convenience, whereas an influencer who occasionally rejects products, discourages purchases, or challenges prevailing preferences demonstrates at least some willingness to incur the social or commercial costs associated with disagreement. The first instances of such disagreement may therefore possess considerable diagnostic value, since they provide evidence against the hypothesis that positive judgments are produced primarily by promotional incentives. However, this same logic does not imply that increasing negativity should produce indefinitely increasing credibility. Once negative judgment becomes sufficiently frequent, it may cease to signify selectivity and instead support alternative interpretations, including chronic dissatisfaction, cynicism, strategic differentiation, performative skepticism, or contrarianism, each of which alters the meaning of negativity from evidence of independent judgment into evidence of a stable disposition or calculated public persona.

The central hypothesis developed here is therefore that the relationship between the frequency of deinfluencing and perceived credibility is likely to be nonlinear, and more specifically that it should, under plausible social assumptions, approximate an inverted-U or otherwise single-peaked function. At low levels of negativity, the marginal credibility benefit of additional negative judgments is expected to be positive because these judgments increase the perceived probability that the influencer exercises independent evaluation rather than simply reproducing promotional incentives. As negativity becomes more common, however, the informational value of each additional negative judgment should diminish, because the audience has already received sufficient evidence that the influencer is willing to disagree, while the reputational costs associated with frequent negativity may begin to increase as the behaviour becomes more compatible with contrarian or strategically differentiated identities.

This formulation also requires a distinction between different forms of trust; an influencer may be regarded as honest while simultaneously being regarded as mistaken, poorly informed, aesthetically incompatible, or socially distant from the observer, and this becomes particularly important when negative evaluations concern products, brands, or practices with which the audience is already positively aligned. Specifically, confirmation bias, prior preference, and identity attachment can alter the interpretation (or functional impact)

of deinfluencing without necessarily eliminating its credibility effect. A negative judgment directed toward a product for which the observer has little prior commitment may be processed primarily as useful information, whereas a negative judgment directed toward an object that is strongly associated with taste, status, community, or self-conception may be experienced not simply as a criticism of the product but as an implicit criticism of the observer's own judgment. Under such conditions, the same act of deinfluencing may increase perceived honesty while reducing perceived expertise, affinity, or benevolence, suggesting that the social consequences of negativity operate in more than one direction and cannot be represented adequately by a single undifferentiated measure of trust.

A universal optimal percentage of negative content is likely to be difficult to predict, as such a value would depend on empirical parameters that are likely to vary across audiences, categories, and influencer roles. The article proposes a formal structure for understanding why an interior optimum should exist at all, why its location should shift across social contexts, and why the point at which negativity maximizes perceived honesty may differ from the point at which it maximizes overall persuasive influence.

2. Inference and Interpretation Mechanisms

The effect of deinfluencing on credibility depends on an interpretive process in which audiences infer latent motives, dispositions, and evaluative standards from observable communication. Influencer communication is useful to consider as a problem of imperfect inference: the audience observes a sequence of recommendations, rejections, qualifications, and endorsements, while the underlying causes of those judgments remain unavailable for direct inspection. A positive recommendation may reflect genuine enthusiasm, commercial incentive, social affiliation, habitual positivity, or some combination of these, and a negative recommendation may similarly arise from independent evaluation, disappointment, status competition, strategic differentiation, or a cultivated contrarian identity. The audience must consequently estimate which explanation is most plausible from the pattern of judgments that it encounters over time.

This process gives individual evaluations meaning beyond their immediate informational content. A recommendation does communicate that a product is desirable, but at the same time it also provides evidence about the type of communicator producing the recommendation. Repeated positivity can gradually alter the interpretation of subsequent positive evaluations because the audience may become less certain that endorsement reflects selectivity. If nearly every product receives approval, the informational distinction between a product that genuinely exceeds the influencer's standards and one that merely appears within a promotional context becomes increasingly difficult to sustain. A negative judgment has greater inferential significance under these circumstances because it demonstrates that endorsement is not automatic, thereby supplying information about the existence of an evaluative threshold.

This argument can initially be expressed using a simple model of latent types. Suppose that an observer considers two broad explanations for an influencer's behaviour. Let I represent an independent evaluator whose recommendations arise primarily from product judgment, and let P represent a promoter whose evaluations are strongly affected by commercial incentives. Before observing any particular review, the audience holds prior beliefs $\Pr(I)$ and $\Pr(P)$. After observing a negative evaluation N, the probability that the influencer is an independent evaluator becomes

$$\Pr(I \mid N) = \frac{\Pr(N \mid I)\Pr(I)}{\Pr(N \mid I)\Pr(I) + \Pr(N \mid P)\Pr(P)}. \tag{1}$$

Whenever

$$\Pr(N \mid I) > \Pr(N \mid P), \tag{2}$$

a negative evaluation increases the posterior probability assigned to independence. The magnitude of that update depends on the relative likelihood of negativity under the competing explanations. If commercially motivated promoters are believed to issue negative recommendations only rarely, then even a small amount of observed negativity can produce a substantial revision of the audience's interpretation of the influencer.

The importance of this formulation lies less in the assumption that consumers literally perform Bayesian calculations than in the structure of the inference itself. Social judgment frequently operates through approximate

probabilistic reasoning in which observed behaviour is assessed according to how compatible it appears with alternative explanations. Audiences may therefore react to negativity by reducing the plausibility of one attribution and increasing the plausibility of another, even when this process is intuitive, affective, or only partially conscious. The relevant question concerns the relative explanatory fit of different social types, since the same observed behaviour can acquire different meanings depending on which latent type it most strongly implies.

A two-type model is useful for explaining why initial negativity can increase credibility, although it becomes insufficient once negativity itself becomes frequent. The interpretation problem then expands to include at least a third type, which can be described as the strategic or dispositional contrarian. Let C denote an influencer whose public identity, engagement strategy, or habitual disposition creates an unusually high probability of negative evaluation. The audience now distributes belief across three explanations,

$$\Pr(I \mid n), \qquad \Pr(P \mid n), \qquad \Pr(C \mid n), \tag{3}$$

where n represents the observed frequency of negative judgments across some relevant period or sequence of content. The critical implication is that increasing n need not continuously increase $\Pr(I \mid n)$. Low values of n may be most compatible with the promotional type, moderate values may be most compatible with the independent type, and very high values may become increasingly compatible with the contrarian type.

The interpretation of negativity can therefore be understood as a movement across attributional regions. When negative judgments are rare, their appearance may provide strong evidence that the influencer is capable of resisting promotional pressure and applying a genuine evaluative standard. When they become more common, their additional informational value declines because the audience has already established that the influencer is willing to reject products. At still higher frequencies, the same behaviour begins to support a different explanation in which negativity reflects a stable communicative style, an audience-positioning strategy, a source of engagement, or a broader disposition toward skepticism. The meaning of negativity changes as its frequency changes, since the observer is continually evaluating which latent

motive offers the most coherent explanation of the behavioural pattern as a whole.

This mechanism also introduces an important distinction between independence and unpredictability. A communicator does not necessarily appear independent merely because their evaluations are difficult to anticipate, since randomness itself provides little evidence of coherent judgment. What matters is the perception that evaluations vary in accordance with a stable standard that is not reducible to external promotional incentives. Occasional negativity can strengthen this perception because it implies that the influencer possesses criteria capable of generating both acceptance and rejection. Excessive negativity may weaken the same perception if the audience comes to believe that rejection has become the default outcome regardless of product quality. In this sense, credibility depends partly on whether positive and negative judgments appear to arise from discriminating evaluation rather than from a fixed directional tendency.

The audience's prior beliefs also have an effect on these inferences. An influencer operating in a highly commercialized category may require more negative evidence before audiences revise an initial expectation of promotional motivation, whereas an influencer with an established reputation for critical independence may require comparatively little. Similarly, the expected base rate of negativity differs across social roles. A professional critic, investigative journalist, luxury reviewer, lifestyle creator, technology commentator, or entertainment influencer enters the interaction with different normative expectations concerning how frequently criticism should occur—one negative-review rate may therefore imply independence in one role and habitual antagonism in another.

The mechanism is consequently sociological as well as cognitive, because the audience is interpreting behaviour through culturally available categories of motive and role. Deinfluencing produces an effect when it shifts the influencer from one of these categories toward another, and the direction of that shift depends on the existing social position of the influencer, the norms attached to the relevant category, and the audience's prior expectations concerning how a credible person in that role ought to behave.

This also explains why the effect cannot be represented adequately by treating each negative review as an independent and equally valuable contribution to trust. The informational value of a judgment depends on the sequence that precedes it. The first negative evaluation observed after a long series of positive endorsements may substantially revise beliefs about the influencer's independence, while another negative evaluation within an already highly critical pattern may add very little evidence of independence and may instead strengthen the alternative interpretation of habitual negativity. The credibility consequence of a negative judgment is conditional on the accumulated history of prior judgments, i.e., the relevant explanatory variable is the position of negativity within a broader distribution of evaluative behaviour.

3. Competing Interpretations and the Nonlinear Credibility Effect

The inferential mechanism described above implies that negativity should have two opposing effects on credibility, both of which arise from the audience's attempt to determine what repeated negative evaluation reveals about the influencer. At relatively low levels, negative judgments can strengthen the perception of independence because they demonstrate that endorsement is not automatic and that some evaluative threshold appears to govern the influencer's recommendations. As negativity becomes increasingly frequent, however, the same behavior can support a different interpretation in which criticism reflects a stable tendency toward rejection, strategic differentiation, or a cultivated contrarian identity. The expected relationship between negativity and credibility therefore depends on the changing balance between these interpretations rather than on negativity having a uniformly positive or negative effect.

Let $n \in [0, 1]$ denote the proportion of an influencer's evaluative content that is negative within some relevant observation period. The first component of the model represents the credibility gained through perceived independence. This benefit should increase with n, although its marginal value is unlikely to remain constant. The first few negative judgments provide relatively strong evidence that the influencer is capable of rejecting products and is therefore

not simply reproducing promotional incentives, whereas later negative judgments add progressively less information once independence has already been established. A convenient representation of this process is

$$I(n) = A\big(1 - e^{-kn}\big), \tag{4}$$

where $A > 0$ represents the upper limit of credibility that can be gained through independence signaling and $k > 0$ represents the rate at which audiences update their interpretation when negative evaluations begin to appear.

The marginal effect of negativity on perceived independence is then

$$I'(n) = Ak\, e^{-kn}, \tag{5}$$

which is positive for all n, while

$$I''(n) = -Ak^2 e^{-kn} < 0. \tag{6}$$

The substantive implication is that the credibility benefit of negativity exhibits diminishing returns. Moving from an entirely positive evaluation history to one containing occasional rejection may meaningfully change the audience's understanding of the influencer, because it establishes the existence of an evaluative boundary. Moving from an already mixed pattern of evaluations to a somewhat more negative pattern may produce a smaller revision, since the relevant inference—that the influencer is willing to disagree—has already been substantially supported.

The same increase in n also changes the plausibility of competing social explanations. As negative judgments become common, the audience may begin to infer that rejection itself is a relatively stable feature of the influencer's behavior. This may be interpreted as excessive skepticism, dispositional negativity, a preference for adversarial positioning, or a strategic attempt to gain attention and distinctiveness through criticism. These interpretations introduce a reputational cost that is initially weak, because occasional negativity is not sufficient to support them strongly, but which may increase more rapidly once criticism becomes habitual.

This second process can be represented as

$$C(n) = Bn^p, \tag{7}$$

where $B > 0$ captures the general sensitivity of credibility to excessive negativity and $p > 1$ allows the cost to accelerate as negative behavior becomes more frequent. Its marginal effect is

$$C'(n) = Bp\, n^{p-1}. \tag{8}$$

At low values of n, this term remains relatively small, because sparse criticism offers little basis for concluding that negativity is itself the influencer's dominant orientation. At higher values, each additional negative judgment becomes more consistent with the contrarian interpretation and therefore contributes more strongly to the reputational cost.

The two components can be combined into a single credibility function,

$$T(n) = T_0 + A\big(1 - e^{-kn}\big) - Bn^p, \tag{9}$$

where T_0 represents baseline credibility prior to considering the observed distribution of positive and negative evaluations. The derivative is

$$T'(n) = Ak\, e^{-kn} - Bp\, n^{p-1}. \tag{10}$$

This expression captures the central theoretical claim of the article. The effect of additional negativity depends on whether the marginal independence benefit exceeds the marginal contrarianism cost. At low levels of n, the first term is likely to dominate, so an increase in negativity raises credibility. As n increases, the independence term declines while the negativity cost increases, reducing the size of the credibility gain and eventually reversing its direction. The trust-maximizing level of negativity, denoted n^*, is therefore defined implicitly by

$$Ak\, e^{-kn^*} = Bp\, (n^*)^{p-1}. \tag{11}$$

At this point, the marginal credibility gained from additional evidence of independence is equal to the marginal credibility lost through increasingly negative attribution.

The significance of n^* is theoretical rather than numerical. No universal value is implied, because the parameters A, k, B, and p depend on the audience, influencer, category, and social context in which evaluation occurs. The model

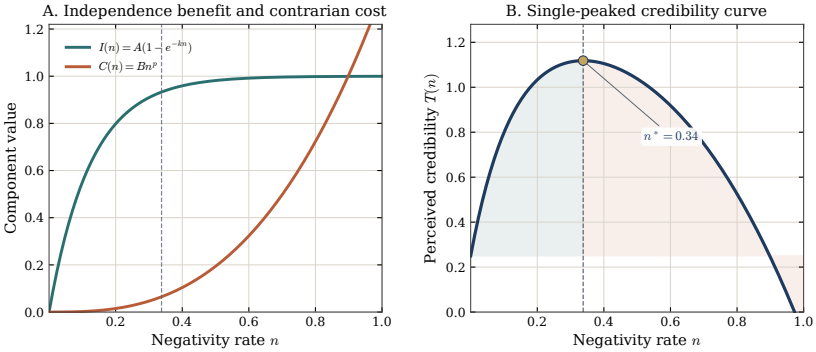

Figure 1: Illustrative credibility function $T(n) = T_0 + I(n) - C(n)$, with $I(n) = A(1 - e^{-kn})$ and $C(n) = Bn^p$. Panel A shows the saturating independence benefit and the accelerating contrarian cost. Panel B plots their combination. The marker is the interior optimum n^*, at which $T'(n^*) = 0$, equivalently $I'(n^*) = C'(n^*)$. Parameterization: $T_0 = 0.25$, $A = 1$, $k = 8$, $B = 1.35$, $p = 2.8$. The numerical location of n^* is illustrative, not a universal recommendation.

instead proposes that, where both processes are present, an interior maximum becomes plausible and empirically estimable.

This formulation also clarifies why the expected nonlinearity is sociologically meaningful. The changing slope of the function corresponds to a changing interpretation of the influencer's behavior. At very low levels of negativity, the dominant question concerns whether the influencer is capable of independent rejection at all. Moderate negativity provides evidence that positive recommendations are selective and therefore potentially more diagnostic. As negative judgments become increasingly common, the relevant interpretive problem changes, because the audience begins to ask whether rejection itself has become predictable. The same observable behavior can therefore move from being evidence against commercial dependence to being evidence for another kind of behavioral regularity.

The distinction can be expressed in terms of competing latent attributions. Let P, I, and C represent promotional, independent, and contrarian interpretations respectively. Increasing negativity is expected to reduce $\Pr(P \mid n)$ initially, increase $\Pr(I \mid n)$ over an intermediate range, and eventually increase $\Pr(C \mid n)$ as the observed rate of rejection becomes more compatible

with systematic negativity. The proposed credibility curve is therefore closely connected to a shift in the relative likelihood assigned to these interpretations:

$$P \longrightarrow I \longrightarrow C. \tag{12}$$

This progression should be understood probabilistically rather than as a set of sharply separated categories. Audiences may assign some probability to all three interpretations simultaneously, and the credibility effect depends on their relative weights. A moderately negative influencer may still be perceived as commercially motivated to some degree, while a highly negative influencer may still be regarded as genuinely independent. What changes with n is the balance among these explanations and, consequently, the meaning attached to additional negative behavior.

There is also an important reflexive element in this process. As deinfluencing becomes more widely recognized as a credibility-enhancing strategy, the inferential value of negativity may itself change. A behavior that initially functions as evidence against commercial motivation can become strategically attractive precisely because audiences treat it as such evidence. Influencers and marketers who understand the credibility value of criticism may therefore have incentives to produce selective negativity for reputational purposes, which introduces a second-order problem in which the audience must consider whether apparently costly dissent is itself being used instrumentally. In formal terms, an increase in the perceived strategic use of deinfluencing would be expected to reduce A, increase B, or alter the rate parameter k, since the same negative judgment would carry less information about genuine independence and more information about deliberate credibility management.

4. The Credibility Curve

The preceding model suggests that the relationship between negativity and credibility can be represented as the interaction of a diminishing independence benefit and an increasing reputational cost, producing the general function

$$T(n) = T_0 + A\big(1 - e^{-kn}\big) - Bn^p, \tag{13}$$

where $n \in [0, 1]$ denotes the observed proportion of negative evaluations, T_0 represents baseline credibility, A describes the maximum credibility that

can be gained through perceived independence, k determines how rapidly this benefit is realized, B represents the strength of the reputational cost associated with frequent negativity, and $p > 1$ determines how sharply that cost accelerates. The value of the function lies primarily in the relationships among these parameters, because the model is intended to describe a family of possible credibility curves whose shapes depend on social conditions.

At the lower end of the distribution, the curve is expected to rise relatively quickly because an initial departure from uniformly positive evaluation carries substantial inferential weight. Where n is close to zero, audiences possess relatively little behavioral evidence that the influencer is willing to reject products, and the introduction of even occasional criticism may therefore produce a considerable change in the perceived probability of independent judgment. This is reflected in the marginal independence term,

$$I'(n) = Ak\, e^{-kn}, \tag{14}$$

which reaches its highest value when negativity is scarce and declines continuously as negativity becomes more common. In sociological terms, the early portion of the curve represents a period in which dissent is unusual enough to remain diagnostic: negative judgment distinguishes the influencer from the behavioral pattern expected of an indiscriminate promoter and consequently changes how subsequent positive recommendations are interpreted.

The importance of this early increase extends beyond the negative evaluations themselves, because the introduction of rejection can retrospectively affect the credibility of positive endorsement. Once an audience has observed that the influencer is capable of saying no, a later positive recommendation can be understood as having passed through a selective process whose alternative outcome was genuine rejection. The value of negativity is partly relational. Its presence changes the meaning of positivity by demonstrating that approval is contingent, and the credibility benefit produced by deinfluencing may consequently appear not only in audience responses to negative content but also in the greater diagnostic value assigned to later endorsements.

As n enters an intermediate range, the curve should begin to flatten because much of the available evidence concerning independence has already been obtained. An audience that has repeatedly observed the influencer rejecting

products no longer requires each subsequent negative judgment to establish that rejection is possible, and additional criticism therefore produces progressively smaller revisions in perceived independence. This region is important because it is likely to contain the point at which the influencer is most clearly interpreted as selective: positive and negative judgments both occur often enough that neither appears predetermined, allowing the pattern as a whole to be attributed to discrimination among alternatives.

The intuitive meaning of this region can be expressed through the concept of evaluative entropy. If the influencer's judgments are almost always positive, the outcome of the next evaluation is highly predictable; if the judgments are almost always negative, the same problem reappears in the opposite direction. Somewhere between these extremes, the audience faces greater uncertainty concerning whether the next product will be accepted or rejected, and this uncertainty can support the inference that the judgment depends on characteristics of the object being evaluated rather than on a fixed directional tendency of the evaluator. The relationship between unpredictability and credibility is not necessarily direct, since random or incoherent judgment would not be expected to generate trust, although a degree of outcome variability is required if audiences are to infer that evaluation is genuinely discriminating.

The curve reaches its maximum at the point n^* satisfying

$$T'(n) = Ak\,e^{-kn} - Bp\,n^{p-1} = 0, \tag{15}$$

or equivalently,

$$Ak\,e^{-kn^*} = Bp\,(n^*)^{p-1}. \tag{16}$$

This equality gives the optimum a more precise interpretation than the informal claim that audiences prefer "some negativity." At n^*, an additional negative judgment would contribute as much credibility through its remaining independence value as it would remove through the increasing plausibility of excessive or strategic negativity. Below this point, further negativity has a positive net effect on credibility; above it, the marginal reputational cost exceeds the remaining informational benefit. The optimum is therefore generated by an equality of marginal social interpretations rather than by an intrinsic preference for moderation.

The position of n^* depends strongly on the parameters of the model. A larger value of A raises the potential return from demonstrating independence and should generally permit a higher level of negativity before the reputational cost becomes dominant. A larger k, by contrast, means that audiences infer independence quickly, so relatively little negativity is required before the informational benefit begins to saturate. A larger B shifts the optimum toward lower values of n, since audiences impose a stronger reputational penalty on frequent criticism, while a larger p produces a cost that remains relatively weak at moderate levels and then rises more sharply once negativity becomes common. The location of the optimum therefore contains information about the social environment in which evaluation occurs.

This dependence is important because there is little reason to assume that two audiences observing the same influencer would necessarily possess the same credibility curve. An audience that begins with strong suspicion of commercial influence may assign substantial value to early negative judgments, increasing A or altering k, whereas an audience already convinced of the influencer's independence may gain comparatively little from further evidence of rejection. An audience accustomed to highly critical commentary may impose only a modest cost on frequent negativity, effectively lowering B, while an audience that expects aspirational or supportive content may penalize the same pattern considerably more strongly. The curve can therefore shift horizontally and vertically without any change in the influencer's actual review distribution.

Baseline credibility T_0 introduces another distinction. Two influencers may possess essentially identical marginal responses to negativity while occupying very different absolute levels of trust because one enters the interaction with an established reputation and the other does not. In the simplest form of the model, T_0 shifts the curve vertically without changing the value of n^*, although this assumption may become unrealistic when prior reputation also alters how audiences interpret subsequent behavior. A highly trusted influencer may require less negative evidence to establish independence, while a distrusted influencer may encounter a more difficult problem in which negative reviews themselves are interpreted as strategic attempts at reputation repair.

In an extended model, T_0 may therefore interact with A, k, or B, making the effect of negativity path-dependent.

The model is similarly sensitive to the time horizon over which n is evaluated. Negativity observed across five recent posts may carry a different meaning from the same numerical proportion observed across several years of content, since audiences do not necessarily weight all historical evaluations equally. A more realistic specification could therefore replace the simple proportion n with a weighted negativity measure,

$$n_t = \frac{\sum_{j=1}^{m} w_j N_j}{\sum_{j=1}^{m} w_j}, \tag{17}$$

where $N_j \in \{0, 1\}$ indicates whether evaluation j was negative and w_j represents the salience or recency of that observation. If recent evaluations receive greater weight, then an influencer can move along the perceived credibility curve without changing their lifetime distribution of positive and negative judgments. This allows the model to incorporate shifts in persona, temporary periods of unusually critical content, and the possibility that audiences update their understanding of an influencer more rapidly than a simple historical average would imply.

The same framework also suggests that the curve may be asymmetric around its maximum. There is no theoretical requirement that being ten percentage points below n^* should have the same credibility consequence as being ten percentage points above it. The independence benefit approaches saturation gradually, while the contrarianism cost may emerge relatively slowly and then accelerate, producing a curve whose rise and decline differ substantially in steepness. In some social settings, audiences may forgive excessive positivity more readily than excessive negativity, while in others a commercially saturated environment may make constant endorsement especially damaging and allow relatively high levels of criticism before credibility deteriorates.

It is also possible that the observed function contains a relatively broad plateau rather than a sharply defined peak. If the marginal independence benefit and marginal negativity cost remain similar across a substantial interval, multiple negativity rates may produce approximately equivalent levels of perceived credibility. This possibility is theoretically useful because social

judgment is unlikely to discriminate perfectly between, for example, an influencer who rejects 25 percent of products and one who rejects 30 percent. In such cases it may be more appropriate to speak of an optimal region of dissent, N^*, rather than a single optimum:

$$N^* = \{n : T(n) \geq T_{\max} - \varepsilon\}, \tag{18}$$

where ε represents some substantively negligible difference in credibility. Such a formulation avoids assigning artificial precision to a social process that may in practice tolerate a range of evaluative patterns.

The curve should also be interpreted dynamically, because successful deinfluencing may alter the conditions that initially made it credible. If audiences increasingly learn that occasional negativity is useful for signaling authenticity or independence, then negative recommendations become less costly as signals and potentially more attractive as strategic behavior. In the language of the model, the credibility benefit associated with independence may weaken as the audience becomes more skeptical about the motive underlying the negative judgment, while the cost associated with strategic contrarianism may emerge at lower levels of n. The credibility curve can therefore evolve culturally: a level of negativity that appeared unusually candid when deinfluencing was uncommon may later appear calculated once the practice becomes established as a recognizable marketing technique.

For this reason, n^* should not be interpreted as a stable behavioral prescription for influencers. It is better understood as an equilibrium generated by current audience expectations regarding promotion, independence, criticism, and strategic self-presentation, all of which can change as communicative practices diffuse. The form of the effect is consequently both an outcome of social interpretation and a potential cause of further social adaptation, because once actors recognize that audiences reward selective dissent, they may alter their behavior in ways that subsequently change the audience's interpretation of dissent itself. The curve is therefore reflexive: social expectations determine its form, behavior responds to those expectations, and repeated strategic behavior can modify the expectations from which the curve originally emerged.

5. There Is No Universal Optimum

The credibility curve implies the existence of an interior optimum under a given set of social conditions, although it does not imply that there is a single optimal level of negativity that applies across influencers, audiences, categories, or periods of time. The value n^* is determined by the parameters governing the credibility benefit of independence and the reputational cost of excessive negativity, and those parameters are themselves products of audience expectations, category norms, prior beliefs, perceived role obligations, and the broader cultural meaning attached to criticism. The model therefore predicts a family of credibility curves rather than one fixed function, with each social context generating a different balance between the value of dissent and the cost of appearing habitually negative.

This can be expressed by allowing the optimum to depend on a vector of contextual variables X,

$$n^* = f(X), \tag{19}$$

where X may contain audience characteristics, category expectations, influencer role, prior reputation, perceived commercialization, and the historical prevalence of deinfluencing within the relevant media environment. In a more explicit form, the underlying credibility function can be written as

$$T(n \mid X) = T_0(X) + A(X)\big(1 - e^{-k(X)n}\big) - B(X)\, n^{p(X)}. \tag{20}$$

The implication is that the same observed negativity rate can occupy very different positions on the credibility curve depending on the context in which it is interpreted. A value of $n = .30$, for example, may fall below the optimum for an audience that expects strong critical scrutiny, close to the optimum for an audience that values selective disagreement, and above the optimum for an audience that interprets repeated criticism as hostile or strategically performative. The observed behavioral rate is therefore insufficient on its own to predict credibility, because its meaning depends on the social expectations against which it is compared.

One important source of variation is the expected role of the communicator. Audiences approach different influencer types with different assumptions about the appropriate distribution of positive and negative judgments. A professional

critic is expected to discriminate strongly among alternatives, and a relatively high rejection rate may therefore provide little evidence of unusual negativity. An aspirational lifestyle influencer, by comparison, may occupy a role in which enthusiasm, recommendation, and positive identification are more strongly expected, causing the same rejection rate to appear unusually critical. These role expectations effectively alter the cost parameter B, since the reputational penalty associated with negativity depends partly on whether the behavior violates the normative expectations attached to the communicator's position.

The expected base rate of criticism within a category should operate similarly. Product categories differ in the extent to which selective rejection appears reasonable or excessive. A consumer may expect a specialist reviewing expensive technical products to reject a substantial proportion of available options, because high standards and detailed comparison are part of the presumed evaluative role. In categories organized more heavily around lifestyle identification, entertainment, or symbolic consumption, repeated rejection may be interpreted less as technical discrimination and more as an expression of taste, temperament, or group alignment. Consequently, the same amount of deinfluencing may produce different credibility effects even when the influencer's underlying motives remain unchanged.

Audience priors create another source of heterogeneity. Suppose two observers encounter the same influencer with different initial beliefs about commercial motivation. One begins with substantial skepticism and assigns a relatively high prior probability to the promotional explanation, while the other already regards the influencer as independent. The first observer has more room to update in response to negative evidence, so an early instance of deinfluencing may generate a comparatively large credibility gain. For the second observer, the same negative judgment provides less new information because independence was already considered likely. In the terms of the model, prior skepticism may affect both the potential magnitude A of the independence benefit and the rate k at which that benefit is realized.

This produces an important distinction between the optimum for an individual and the optimum observed at the population level. If audience members

possess different parameter values, then each observer i can be represented as having an individual credibility function

$$T_i(n) = T_{0i} + A_i\big(1 - e^{-k_i n}\big) - B_i\, n^{p_i}, \tag{21}$$

with an associated optimum

$$n_i^* = \arg\max_n T_i(n). \tag{22}$$

The aggregate response observed in empirical research is then some distribution of these individual curves rather than a single homogeneous response. An apparent population-level optimum may therefore conceal substantial disagreement among audience members, with some observers continuing to gain trust from additional negativity while others have already entered the declining portion of their own credibility curve.

The population average can be represented as

$$\bar{T}(n) = E\big[T_i(n)\big], \tag{23}$$

although maximizing the average,

$$n_{\text{avg}}^* = \arg\max_n \bar{T}(n), \tag{24}$$

does not imply that n_{avg}^* is optimal for any particular audience member. This distinction matters when interpreting empirical estimates, because an aggregate inverted-U may emerge from multiple heterogeneous response functions, and the location of its maximum may provide only limited information about the underlying social process unless the relevant sources of heterogeneity are also modeled.

Cultural context should also affect the shape and location of the curve. Norms concerning politeness, skepticism, criticism, expertise, commercial endorsement, and public disagreement vary across social environments, and these differences influence the meaning assigned to negative evaluation. In settings where skepticism toward advertising is widespread and criticism is associated with sophistication or consumer competence, a higher negativity rate may be tolerated before contrarian attribution becomes dominant. In environments where public criticism is more strongly associated with antagonism,

disloyalty, or interpersonal conflict, the reputational cost may emerge earlier. The value of n^* is partly a cultural variable, even when the psychological mechanisms of inference remain broadly similar.

The same logic applies temporally. The meaning of deinfluencing in a media environment where the practice is unusual should differ from its meaning after audiences become familiar with it as an established content strategy. Early adopters may benefit from a relatively large independence signal because negative recommendations violate the dominant expectation of promotional positivity. As deinfluencing becomes normalized, its novelty and diagnosticity decline, while awareness of its strategic value may increase skepticism concerning the motive behind the behavior. A culturally successful credibility strategy can therefore modify the conditions that made the strategy effective in the first place.

This introduces a reflexive relationship between the credibility curve and influencer behavior. If creators learn that moderate negativity increases trust, they may strategically adjust their content toward the estimated credibility-maximizing region. As this pattern becomes visible, audiences may update their beliefs about the probability that negativity is itself commercially or reputationally motivated. The resulting change in audience interpretation may reduce the value of the original strategy, shifting the curve and producing a new optimum. What appears to be an optimal behavioral rate at one moment may become suboptimal after the practice is widely imitated.

The absence of a universal optimum therefore follows directly from the social nature of the mechanism. The model does not assume that people possess an intrinsic preference for some fixed proportion of negative information. It assumes that they interpret patterns of evaluation in relation to expectations about how different types of communicators are likely to behave, and those expectations vary across people and contexts. The relevant theoretical object is accordingly not a universal value of n^*, but the set of variables that determine its location and the direction in which changes in those variables shift the curve. This distinction also limits the normative interpretation of the model—even if an empirical study were to estimate a particular negativity rate as credibility-maximizing within one sample, category, and period, that estimate should not be treated as a general recommendation for influencer

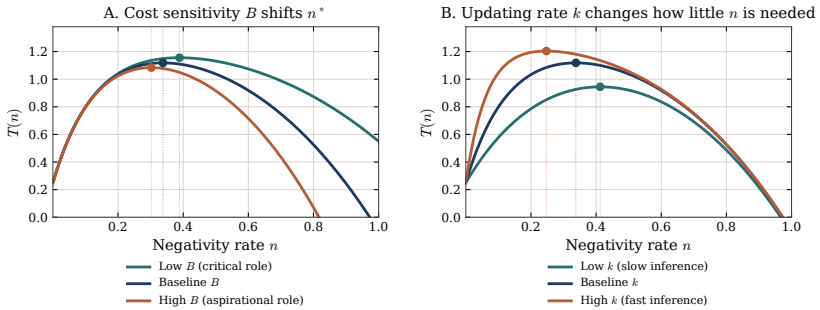

Figure 2: Comparative statics of the credibility curve. Panel A varies the reputational-cost parameter B: a critical or already skeptical audience (low B) tolerates more negativity before the peak; an aspirational role (high B) moves n^* leftward. Panel B varies the updating rate k: when audiences infer independence quickly, little negativity is required before the independence term saturates. Markers indicate each curve's maximum. Values are illustrative.

behavior. Its substantive value would lie in identifying the parameters and conditions that generated the optimum, because those conditions determine whether the result can plausibly transfer elsewhere.

6. Preference Alignment and Confirmation Bias

The credibility curve developed above treats negativity as a property of the influencer's evaluative behavior, yet the effect of any particular negative judgment also depends on the audience's prior relationship to the object being criticized. A negative evaluation directed toward a product for which the observer has little prior preference can be processed primarily as new information, whereas criticism of a product, brand, practice, or cultural object toward which the observer is already favorably disposed creates a more complicated interpretive problem. In such cases, deinfluencing does not merely provide evidence about the influencer; it also conflicts with an existing judgment held by the audience, and the resulting response is shaped by the strength of that prior belief and by the extent to which the object has become connected to the observer's identity, taste, or social position.

Let q_j represent the observer's prior evaluation of object j, and let r_j represent the influencer's evaluation of the same object. A simple measure of disagreement can be written as

$$d_j = |q_j - r_j|. \tag{25}$$

The magnitude of d_j provides an initial indication of evaluative distance, although disagreement alone does not determine its social significance. An observer who mildly prefers a particular household product may encounter a strongly negative review without experiencing substantial threat to their own judgment, while an equally strong disagreement concerning a brand, cultural object, or practice that contributes to self-definition may be interpreted much more personally. It is therefore useful to introduce an identity-salience parameter s_j, where higher values indicate that the object is more closely associated with the observer's sense of taste, status, affiliation, or self-conception. The socially relevant disagreement cost can then be represented provisionally as

$$D_j = \lambda s_j \, d_j^{\gamma}, \tag{26}$$

where $\lambda > 0$ captures sensitivity to disagreement and $\gamma \geq 1$ allows the effect of disagreement to increase nonlinearly as evaluative distance becomes more pronounced.

This formulation is closely related to confirmation bias because prior beliefs influence both the acceptance of new information and the evaluation of the person supplying it. When an influencer's judgment is broadly consistent with the observer's existing preference, the negative or positive recommendation can be assimilated relatively easily into the observer's current understanding. When the judgment is inconsistent with a strongly held preference, the observer faces several possible responses: revising the original belief, discounting the evidence, questioning the influencer's expertise, interpreting the disagreement as evidence of incompatible taste, or attributing the position to some strategic or dispositional motive. Confirmation bias thus operates partly through selective updating of the object-level belief and partly through selective updating of beliefs about the source.

This distinction makes it advantageous to treat credibility as multidimensional within the present section, since disagreement may affect separate

judgments in different directions. At minimum, the audience can distinguish between perceived honesty or integrity H, perceived expertise or quality of judgment E, and perceived benevolence or social affinity B_v. Overall credibility can be represented in simplified form as

$$T = w_H H + w_E E + w_B B_v, \tag{27}$$

where the weights reflect the relative importance of each dimension in a given context. A preference-incongruent act of deinfluencing may increase H, because the influencer appears willing to state a genuine and potentially unpopular judgment, while simultaneously decreasing E, because the observer interprets the disagreement as evidence of inferior judgment, and it may also decrease B_v when the criticism is experienced as socially distancing or hostile. The resulting credibility effect can therefore be ambiguous even when the audience becomes more convinced that the influencer is sincere.

This possibility is important because trust and agreement are often treated too closely in studies of persuasive communication. An observer may regard an influencer as honest while considering their conclusions unreliable, and may therefore become more certain about the influencer's motives while becoming less willing to follow their recommendations. Deinfluencing can thus produce a divergence between moral or motivational credibility and epistemic credibility. The former concerns whether the influencer is believed to communicate sincerely and independently, while the latter concerns whether the influencer is believed to possess sound judgment. Preference alignment is particularly relevant to the second component because persistent disagreement can gradually transform the audience's interpretation from "this person is candid" to "this person's evaluative standards do not correspond to mine."

The interaction becomes stronger when the object being criticized carries symbolic or identity value. Consumer objects frequently serve functions that exceed their immediate utility, including the communication of status, belonging, taste, morality, expertise, or cultural affiliation. Under these conditions, a negative evaluation can be interpreted as indirectly evaluating the people who choose or value the object. The criticism of the product may therefore

be translated into an inference concerning the consumer's own competence or identity:

$$\begin{aligned} &\text{negative evaluation of object} \\ &\quad \longrightarrow \text{ negative implication for chooser} \\ &\quad \longrightarrow \text{ identity-relevant disagreement.} \end{aligned} \tag{28}$$

This process need not be consciously articulated for it to affect the evaluation of the influencer. An observer can experience a reduction in affinity, openness, or perceived expertise simply because the influencer's judgment threatens a preference that has become socially or personally meaningful.

The implications for the credibility curve are substantial. The earlier function,

$$T(n) = T_0 + A\big(1 - e^{-kn}\big) - Bn^p, \tag{29}$$

assumes that negative judgments contribute to credibility according to their frequency and the attributions they generate, yet the effective value of a negative judgment may also depend on whether it confirms or contradicts the audience's prior preferences. A more complete specification might therefore allow the marginal benefit and cost terms to depend on alignment,

$$T(n, a) = T_0 + A(a)\big(1 - e^{-k(a)n}\big) - B(a)\, n^{p(a)}, \tag{30}$$

where a represents audience–influencer preference alignment. High alignment may allow the influencer to sustain a larger amount of negativity without suffering a substantial decline in perceived judgment, because the criticism generally confirms the observer's own evaluative standards. Low alignment may produce the opposite result, causing contrarian or incompetence attributions to arise at lower levels of n.

This gives preference alignment a role in determining the location of the optimum. If $n^*(a)$ denotes the negativity rate that maximizes credibility for an observer with alignment level a, then it is plausible that

$$\frac{\partial n^*}{\partial a} > 0 \tag{31}$$

over at least part of the relevant domain, meaning that greater alignment permits a higher level of observed negativity before credibility begins to decline.

The same logic also produces an asymmetry in confirmation bias. Positive evaluations of objects already favored by the observer are likely to reinforce both product belief and source affinity, while negative evaluations of those objects force a choice between revising the product belief and revising the source evaluation. Where prior attachment is weak, updating the product belief may be relatively inexpensive. Where attachment is strong, discounting the influencer may become psychologically and socially less costly than abandoning the existing preference. The effect of deinfluencing therefore depends partly on where the audience places the burden of error: the product may be judged less favorably, or the influencer may be judged less competent, less similar, or less relevant.

This interaction suggests that the credibility-maximizing level of negativity cannot be estimated independently of the distribution of audience preferences. An influencer whose negative judgments tend to coincide with those of the audience may appear discerning across a relatively broad range of n, while an influencer who repeatedly rejects objects valued by the audience may move into the declining region of the credibility curve much earlier. The same nominal negativity rate can consequently produce different social interpretations because the audience is evaluating both the frequency of dissent and the direction of that dissent relative to its own beliefs.

6.1 Deinfluencing as Symbolic Conflict

Preference disagreement becomes more consequential when the object under evaluation carries symbolic meaning beyond its immediate functional qualities. Consumers frequently use brands, products, aesthetic choices, and consumption practices as markers of taste, affiliation, status, morality, competence, or cultural belonging, and criticism directed toward such objects can therefore be interpreted as extending indirectly toward the people who value them. Under these conditions, deinfluencing may cease to operate solely as marketplace information and begin to function as a form of symbolic conflict.

This process can be represented as an escalation in the meaning assigned to disagreement. A negative judgment initially concerns the object itself, although the audience may subsequently infer an evaluation of the kind of person who chooses that object, particularly when the influencer's criticism invokes

taste, intelligence, morality, social awareness, or group membership. The interpretive sequence can therefore be expressed as

$$\begin{aligned} &\text{negative evaluation of object} \\ &\quad \longrightarrow \text{ negative implication for chooser} \\ &\quad \longrightarrow \text{ identity-relevant disagreement.} \end{aligned} \tag{32}$$

The strength of this transition should depend on the symbolic salience of the object. If s_j represents the degree to which object j is incorporated into the observer's identity or social positioning, then higher values of s_j increase the likelihood that disagreement is interpreted personally rather than informationally. In such cases, the credibility consequences of deinfluencing may become more severe because the observer is no longer deciding only whether the influencer is correct about a product, but also whether the influencer's evaluative framework is compatible with the observer's own social identity.

This helps explain why identical levels of negativity may produce sharply different responses across categories. Criticism of a low-involvement functional purchase may invite relatively detached updating, while criticism of a favored fashion label, technology ecosystem, cultural product, political consumption practice, or status-signaling brand may provoke defensive interpretation even when the influencer is regarded as sincere. The resulting response can preserve perceived honesty while reducing affinity, benevolence, or epistemic trust, particularly where the observer experiences the criticism as implying poor judgment on the part of people like themselves. As the identity relevance of the criticized object increases, the social cost of preference-incongruent deinfluencing should rise, moving the credibility curve downward or moving its optimum toward a lower level of negativity for affected audience segments. The same act of dissent can consequently function as evidence of independence for one observer and as a socially threatening judgment for another, depending on whether the criticized object is treated as a neutral consumption choice or as part of the observer's symbolic self-definition.

7. A Bayesian Model with Three Influencer Types

The preceding sections describe the credibility effect as the product of competing social interpretations, and a Bayesian framework provides a useful way of formalizing how those interpretations may change as the observed frequency of negativity increases. The purpose of the model is not to claim that audiences perform explicit probabilistic calculations when evaluating influencers, but to represent the inferential structure of the problem: observers encounter a sequence of positive and negative judgments and gradually update their beliefs about what type of communicator could plausibly have produced that sequence.

A two-type model, distinguishing an independent evaluator from a commercially motivated promoter, is sufficient to explain why initial acts of deinfluencing may increase perceived credibility. It becomes less satisfactory once negativity itself becomes common, because repeated rejection can then support a qualitatively different attribution. The model therefore introduces three latent influencer types:

$$P = \text{promotional type}, \tag{33}$$

$$I = \text{independent evaluative type}, \tag{34}$$

$$C = \text{contrarian type}. \tag{35}$$

The promotional type is characterized by a relatively low probability of issuing negative judgments, because its communication is assumed to be strongly shaped by commercial or reputational incentives favoring endorsement. The independent type is assumed to generate both positive and negative judgments according to some substantive evaluative standard, producing an intermediate rate of negativity. The contrarian type is characterized by a comparatively high probability of negative judgment, whether because criticism has become a stable disposition, an identity position, an engagement strategy, or a means of differentiation from mainstream promotional content.

Let

$$\theta_P,\ \theta_I,\ \theta_C \tag{36}$$

denote the expected probability of a negative judgment for each type, with the ordering

$$\theta_P < \theta_I < \theta_C. \tag{37}$$

This ordering is central to the model, because it allows the same observed behavior to carry different inferential implications depending on its frequency. A low rate of negativity is most compatible with the promotional type, a moderate rate is most compatible with the independent type, and a high rate increasingly supports the contrarian type. The model does not require the three types to be mutually exclusive in any psychological sense; they are better interpreted as stylized explanatory categories that audiences may use when making sense of repeated evaluative behavior.

Suppose that an observer has seen m evaluations from an influencer, of which x are negative. The observed negativity rate is

$$n = \frac{x}{m}. \tag{38}$$

Conditional on type $z \in \{P, I, C\}$, the simplest specification assumes that the number of negative evaluations follows a binomial distribution,

$$x \mid z \sim \text{Binomial}(m, \theta_z). \tag{39}$$

The likelihood of observing x negative evaluations from a given type is therefore

$$\Pr(x \mid z) = \binom{m}{x} \theta_z^x (1 - \theta_z)^{m-x}. \tag{40}$$

Given prior beliefs

$$\Pr(P), \qquad \Pr(I), \qquad \Pr(C), \tag{41}$$

the posterior probability assigned to type z becomes

$$\Pr(z \mid x) = \frac{\Pr(x \mid z)\Pr(z)}{\sum_{r \in \{P,I,C\}} \Pr(x \mid r)\Pr(r)}. \tag{42}$$

The combinatorial term $\binom{m}{x}$ is common to all three likelihoods and therefore cancels when the observer compares the relative plausibility of the competing

types. The posterior probabilities are determined by the extent to which the observed rate of negativity resembles the rate expected under each attribution, weighted by the audience's prior beliefs.

The substantive implication is that the posterior probability of independence need not increase monotonically with negativity. Beginning from a very low value of n, an increase in observed negative judgments may substantially reduce the probability assigned to the promotional explanation, because such behavior is relatively unlikely under θ_P. Over some intermediate range, the observed sequence becomes increasingly compatible with the independent type, causing

$$\Pr(I \mid x) \tag{43}$$

to rise. Once the negativity rate moves sufficiently far above the level expected from an independent evaluator, however, additional negative judgments become increasingly compatible with the contrarian type, and

$$\Pr(C \mid x) \tag{44}$$

begins to dominate. Figure **??** shows how those posteriors, and a credibility index built from them, move with n.

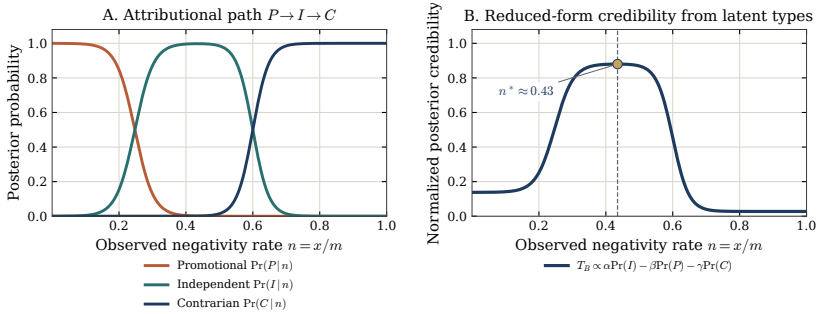

Figure 3: Three-type inference. Panel A shows posterior probabilities of the promotional (P), independent (I), and contrarian (C) types as the observed negativity rate $n = x/m$ rises ($\theta_P = 0.12$, $\theta_I = 0.38$, $\theta_C = 0.78$; $m = 24$; priors 0.50, 0.30, 0.20). Panel B plots a normalized credibility index $T_B = \alpha \Pr(I \mid x) - \beta \Pr(P \mid x) - \gamma \Pr(C \mid x)$. The reduced-form inverted U in Figure **??** approximates this attributional competition. The plotted n^* is an illustration for these parameters, not a general estimate.

The Bayesian model therefore produces the attributional progression proposed earlier:

$$P \longrightarrow I \longrightarrow C. \tag{45}$$

The important feature of this progression is that it arises from relative likelihood rather than from any inherent positive or negative meaning attached to criticism. Negativity increases credibility while it makes the independent explanation more plausible relative to its alternatives. Its credibility value declines once further negativity makes another explanation more plausible than independence. This gives the nonlinearity a probabilistic foundation: the same observable variable changes meaning because it moves the audience through different regions of the latent-type distribution.

The transition points between these regions can be derived directly from the model. Consider first the point at which the independent type becomes more probable than the promotional type. Ignoring the common binomial coefficient, the observer prefers I to P when

$$\Pr(x \mid I)\Pr(I) > \Pr(x \mid P)\Pr(P). \tag{46}$$

Taking logarithms gives

$$x \log\Big(\frac{\theta_I}{\theta_P}\Big) + (m - x)\log\Big(\frac{1-\theta_I}{1-\theta_P}\Big) + \log\Big(\frac{\Pr(I)}{\Pr(P)}\Big) > 0. \tag{47}$$

Since $x = mn$, this can be written in terms of the observed negativity rate as

$$mn \log\Big(\frac{\theta_I}{\theta_P}\Big) + m(1-n)\log\Big(\frac{1-\theta_I}{1-\theta_P}\Big) + \log\Big(\frac{\Pr(I)}{\Pr(P)}\Big) > 0. \tag{48}$$

An equivalent expression can be written for the crossover between the independent and contrarian types. These boundaries are useful because they formalize the idea that the same negativity rate can be interpreted differently under different priors. If an audience begins with a strong expectation that influencers are commercially motivated, the amount of negative evidence required to move from P toward I will differ from that required by an audience that already regards the communicator as relatively independent. Likewise, if the social environment makes strategic contrarianism appear common, the transition from I toward C can occur at a lower negativity rate.

The sample size m also matters. A negativity rate of $n = .50$ based on two observed reviews carries considerably less inferential weight than the same rate observed across one hundred reviews, because the larger sequence provides more evidence that the pattern reflects a stable behavioral tendency. The posterior distribution should therefore become more concentrated as m increases. This feature is important for influencer contexts, where audiences differ substantially in how much of a creator's history they have observed. A new viewer and a long-term follower may encounter the same current review while holding very different posterior beliefs because one evaluates it against a sparse (or no) history and the other against a much more extensive behavioral record.

This introduces a distinction between the observed negativity rate and confidence in the attribution derived from that rate. Two influencers can exhibit the same value of n, while the social certainty surrounding their perceived type differs according to the quantity and consistency of prior observations. In a fuller model, credibility may therefore depend on both the posterior probability of independence and the concentration of the posterior distribution. An audience that assigns probabilities of 0.40, 0.35, and 0.25 to the three types is considerably less certain about the influencer than an audience assigning 0.05, 0.90, and 0.05, even though the independent type is the modal interpretation in both cases.

One possible formalization is to let credibility depend on the posterior probability of the independent type,

$$T_B = \alpha \, \Pr(I \mid x), \tag{49}$$

where $\alpha > 0$ scales the contribution of inferred independence to perceived credibility. A richer specification can incorporate penalties associated with the competing interpretations,

$$T_B = \alpha \, \Pr(I \mid x) - \beta \, \Pr(P \mid x) - \gamma \, \Pr(C \mid x), \tag{50}$$

with $\beta, \gamma > 0$. This representation allows promotional and contrarian attributions to reduce credibility by different amounts, which is sociologically plausible because audiences may judge commercial dependence and habitual

contrarianism as distinct forms of unreliability. A promotional influencer may be considered socially pleasant yet motivationally compromised, while a contrarian influencer may be considered independent in motive but unreliable in judgment. The weights β and γ can therefore differ across audiences and contexts.

The Bayesian formulation also clarifies the relationship between the latent-type model and the continuous credibility curve developed earlier. The function

$$T(n) = T_0 + A\left(1 - e^{-kn}\right) - Bn^p \tag{51}$$

can be understood as a reduced-form approximation to a more detailed inferential process in which the posterior probability of independence first rises and later falls as the observed negativity rate moves from the region associated with promotion toward the region associated with contrarianism. The exponential independence term captures the diminishing informational benefit of evidence against the promotional explanation, while the accelerating cost term captures the increasing plausibility of the contrarian explanation. The earlier credibility function therefore summarizes, in continuous form, the same attributional competition represented explicitly by the three-type Bayesian model.

The three-type structure also provides a useful way to incorporate confirmation bias and preference alignment. The expected negativity rate associated with the independent type need not be interpreted identically by all observers, because disagreement with the audience may alter the perceived likelihood that the behavior reflects good-faith evaluation rather than poor judgment or oppositional disposition. If an influencer repeatedly criticizes products that the observer already dislikes, the audience may assign a relatively high likelihood to the independent type, because the evaluations appear consistent with its own standards. If the same influencer repeatedly criticizes highly valued objects, the observer may revise the perceived parameters of the latent types, effectively treating such behavior as more consistent with contrarianism. The likelihood functions are therefore partly socially constructed: the observer is estimating not only how often each type would be negative, but how each type would be expected to evaluate particular objects.

This can be represented by allowing the type-specific negativity probabilities to depend on preference alignment a,

$$\theta_z = \theta_z(a). \tag{52}$$

The posterior then becomes

$$\Pr(z \mid x, a) = \frac{\Pr(x \mid z, a)\Pr(z \mid a)}{\sum_r \Pr(x \mid r, a)\Pr(r \mid a)}. \tag{53}$$

Such an extension makes explicit that inference about influencer type is inseparable from the observer's own evaluative position. The same sequence of judgments may generate different posterior beliefs for different audience members because they possess different priors and different assumptions about what independent, promotional, or contrarian behavior would look like in relation to the objects being discussed.

A further implication concerns strategic adaptation. The model initially assumes that the latent types generate different characteristic rates of negativity, but those rates need not remain stable once creators understand how audiences interpret them. A promotional communicator who learns that occasional negative judgments increase the posterior probability of independence may strategically increase θ_P, thereby making the promotional type behaviorally more similar to the independent type. As the distributions overlap more strongly, negative judgments become less diagnostic. In probabilistic terms, the likelihood ratio

$$\frac{\Pr(N \mid I)}{\Pr(N \mid P)} \tag{54}$$

moves closer to one, reducing the amount of information contained in an observed negative recommendation.

The same process may occur at the upper end of the distribution. A creator who develops a successful identity around skepticism may strategically increase negativity because criticism itself attracts attention, producing behavior that becomes increasingly difficult to distinguish from genuine high-selectivity evaluation. The three latent categories should therefore be understood as historically contingent distributions whose separation can change as influencer practices evolve. If deinfluencing becomes widely adopted as a credibility

strategy, the distinction between P and I narrows, while the boundary between I and C may also become increasingly difficult for audiences to determine.

This dynamic quality is central to the sociological interpretation of the Bayesian model. The observer is not estimating fixed natural types; the observer is interpreting actors who are themselves capable of anticipating how they will be interpreted. Influencers can modify their behavior in response to audience expectations, audiences can update their priors in response to widespread strategic behavior, and the informational value of negativity can consequently change over time. The credibility signal is therefore reflexive: its effectiveness influences its adoption, its adoption influences its diagnosticity, and changes in diagnosticity alter the posterior beliefs through which the signal originally generated credibility.

The three-type model consequently provides a probabilistic foundation for the broader theory of optimal dissent. Low negativity is comparatively compatible with promotional dependence, moderate negativity with selective independence, and high negativity with strategic or dispositional contrarianism, while the exact boundaries between these regions depend on priors, social role, audience alignment, observation history, and the historical prevalence of deinfluencing as a communication strategy. The central theoretical claim is therefore not that a fixed rate of criticism uniquely identifies a credible communicator, but that credibility emerges from the relative plausibility of competing explanations for observed evaluative behavior, and that the probability assigned to those explanations changes systematically as the pattern of negativity changes.

8. Effects of Strategic Contrarianism

The credibility value of deinfluencing creates an incentive for that behavior to be adopted strategically. Once influencers, agencies, and audiences become aware that selective negativity can signal independence, criticism is no longer interpretable only as evidence of genuine evaluative resistance. It also becomes a possible technique of credibility management. This introduces a second-order inference problem in which the audience must evaluate not only whether

the influencer is willing to criticize products, but whether that willingness itself has been incorporated into a broader strategy of self-presentation.

Strategic contrarianism can be understood as the deliberate use of disagreement, rejection, or skepticism in order to produce a desired social interpretation. In influencer contexts, this may involve criticizing some products to strengthen the credibility of later endorsements, adopting a generally skeptical persona in order to differentiate oneself from conventionally promotional creators, or selecting highly visible targets for criticism because negative positioning attracts attention and reinforces an identity built around independence. The resulting behavior may resemble genuine critical evaluation closely enough that the audience cannot infer motive directly from the negative judgment itself.

Within the three-type Bayesian model, the development of strategic contrarianism changes the expected behavioral distributions of the latent types. If the promotional type P learns that negative evaluations can increase perceived independence, its expected negativity rate may increase from θ_P to some strategically adjusted value θ'_P, such that

$$\theta'_P > \theta_P. \tag{55}$$

As θ'_P approaches θ_I, where θ_I denotes the expected negativity rate of an independent evaluator, the behavioral distinction between promotion and independence becomes weaker. The diagnostic value of a negative judgment can be expressed through the likelihood ratio

$$LR_N = \frac{\Pr(N \mid I)}{\Pr(N \mid P)}. \tag{56}$$

If promotional communicators increasingly use negative evaluations, then $\Pr(N \mid P)$ rises and LR_N falls. A negative judgment consequently provides less evidence in favor of the independent interpretation than it would in an environment where commercially motivated influencers almost never criticize products.

The informational consequences extend beyond individual negative judgments. If audiences become familiar with the strategic use of deinfluencing, their prior beliefs may also change. Let S represent the audience's belief that

observed criticism is strategically motivated. Increased cultural awareness of deinfluencing as a credibility technique raises $\Pr(S)$, which should reduce the extent to which negativity is automatically interpreted as evidence of independence. The same observed behavior can therefore generate a smaller posterior credibility update at a later point in time than it generated when the behavior was less common or less widely understood.

This process can be incorporated into the earlier credibility function by allowing the independence benefit to depend on perceived strategic prevalence. Let $s \in [0, 1]$ denote the perceived probability that deinfluencing is strategically deployed within the relevant environment. The maximum credibility benefit associated with independence can then be written as

$$A = A(s), \tag{57}$$

with the plausible condition

$$\frac{\partial A}{\partial s} < 0. \tag{58}$$

As strategic awareness increases, the same amount of negativity produces a smaller potential credibility gain. The rate of initial updating may also change,

$$k = k(s), \tag{59}$$

because audiences who have learned that criticism can be manufactured as a trust signal may require a longer or more consistent history of negative evaluation before revising their beliefs about the influencer's motives.

Strategic contrarianism can also affect the cost side of the credibility function. If audiences associate frequent criticism with attention seeking or reputation management, the contrarianism cost may emerge at lower levels of negativity. The parameter B may therefore increase with perceived strategic prevalence,

$$\frac{\partial B}{\partial s} > 0, \tag{60}$$

producing a credibility curve in which the benefits of initial negativity are weaker and the costs of repeated negativity appear sooner. Under these conditions, the trust-maximizing level n^* may shift toward a lower value even if the influencer's actual evaluative standards have not changed.

The broader implication is that credibility signals can become less reliable through successful imitation. A behavior initially acquires informational value because it is relatively uncommon among actors with strong promotional incentives. Once that value becomes recognized, the behavior attracts imitation from precisely those actors it previously helped audiences distinguish. The result is a partial pooling process in which independent and strategic communicators become more behaviorally similar, forcing audiences to search for additional cues that can separate them.

This dynamic may produce escalating demands for stronger evidence of sincerity. Audiences may begin to evaluate the content of criticism more closely, asking whether the negative judgment appears costly, specific, consistent with prior standards, or directed toward products from which the influencer could plausibly have benefited. A minor criticism of an irrelevant product may carry little evidential weight if it can easily be interpreted as symbolic dissent chosen to protect credibility, while criticism that involves a meaningful commercial sacrifice may remain more diagnostic. The credibility of negativity may therefore depend increasingly on its perceived cost.

A simple extension can capture this possibility. Let c_j denote the perceived cost incurred by issuing negative judgment j. The diagnostic value of the judgment may be written as

$$D_j = f(c_j), \qquad f'(c_j) > 0, \tag{61}$$

so that criticism becomes more informative when the audience believes the influencer has sacrificed a meaningful opportunity by expressing it. A negative review of a product with no sponsorship potential may provide weaker evidence of independence than rejection of a lucrative partnership or a product category closely associated with the influencer's commercial identity. This introduces a distinction between negativity frequency and negativity cost, since two influencers can produce the same proportion of negative content while providing very different amounts of evidence concerning their independence.

Strategic contrarianism also creates incentives at the upper end of the negativity distribution. Some influencers may obtain value directly from adopting

a consistently adversarial position, particularly in environments where criticism attracts engagement, distinguishes the creator from mainstream promotional culture, or provides a recognizable identity around which audiences can organize. For these communicators, negative evaluation can become a form of branding. The probability of criticism then depends partly on its identity and engagement value rather than on the merits of the object under review.

This creates an additional problem for audience inference. A highly negative creator may genuinely be independent of conventional commercial incentives while still being strategically dependent on negativity as a source of attention, social position, or audience loyalty. Independence from direct sponsorship does not imply independence from all incentives. The relevant latent motive may shift from

"I praise this because I am paid to promote it" (62)

to

"I criticize this because criticism sustains my position." (63)

Both forms introduce systematic pressures on evaluation, although they operate in opposite directions.

This possibility suggests that the model should distinguish directional independence from motivational independence. An influencer can be directionally negative while remaining strategically constrained by the rewards associated with negativity. Genuine evaluative independence requires that positive and negative judgments remain responsive primarily to the characteristics of the object and the influencer's stated standards. Strategic contrarianism weakens that relationship by making the direction of the judgment partially responsive to the social rewards of dissent.

The effect is therefore likely to alter the meaning of the credibility curve over time. When deinfluencing is uncommon, moderate negativity may occupy a relatively clear region associated with independence. As strategic adoption increases, the distributions corresponding to promotional, independent, and contrarian types begin to overlap more substantially, reducing the precision with which audiences can infer motive from frequency alone. The credibility curve may flatten, its maximum may become less pronounced, and variables

such as consistency, perceived sacrifice, specificity of criticism, and historical behavior may become increasingly important in determining whether negativity is interpreted as genuine evaluation or strategic performance.

Strategic contrarianism consequently introduces a reflexive process into the theory. Audiences reward selective criticism because it appears to reveal independence; influencers observe this reward and gain incentives to imitate the behavior; audiences gradually recognize the possibility of imitation and revise their interpretation of criticism; the credibility value of negativity then changes, altering the incentives facing influencers once again. The social meaning of deinfluencing is therefore partly endogenous to its own success, and any empirical estimate of an optimal negativity level should be treated as historically contingent on the degree to which audiences currently regard criticism as a credible signal of independent judgment.

9. Discussion

The argument developed in this article begins from a relatively simple observation: negative evaluation can increase credibility because it provides evidence that an influencer is willing to depart from promotional incentives, although the value of that evidence should decline as negativity becomes more common and may eventually reverse once criticism becomes sufficiently frequent to support alternative attributions. The resulting relationship is therefore expected to be nonlinear, with credibility determined by the changing balance between independence signaling and the reputational costs associated with habitual or strategic negativity. The mathematical representation of this process is intended to formalize that balance, while the sociological interpretation explains why the parameters governing the curve should vary across audiences, categories, roles, and historical contexts.

The central contribution is consequently not the claim that some moderate amount of negativity is intuitively preferable to either universal praise or universal criticism. That proposition is relatively easy to anticipate. The more consequential claim concerns the mechanism and graphable form of the relationship: the effect of negativity changes because the audience is revising its interpretation of what repeated evaluative behavior signifies. At low levels,

criticism is beneficial partly because it reduces the plausibility of indiscriminate promotion. At intermediate levels, the influencer may appear most clearly selective, since both approval and rejection remain plausible outcomes. At high levels, criticism becomes increasingly compatible with explanations based on cynicism, contrarianism, strategic differentiation, or identity positioning. The credibility curve therefore represents a movement between competing social attributions rather than a direct psychological preference for moderation itself.

This interpretation also places limits on the meaning of the theoretical optimum n^*. The model does not imply that there exists a universal percentage of negative content that would maximize credibility across influencer contexts. The location of the maximum depends on the parameters that describe audience sensitivity to independence, the speed with which independence is inferred, the social cost associated with frequent criticism, and the degree to which that cost accelerates. These parameters are affected by prior reputation, expected role, category norms, audience composition, cultural attitudes toward criticism, and the prevalence of deinfluencing as an established communication strategy. The relevant object of study is therefore the structure that produces an optimum and the variables that move it, rather than the identification of a context-free numerical value.

A further distinction is necessary between the amount of negativity that maximizes credibility and the amount that maximizes influence. These two outcomes are related, although there is little reason to expect their maxima to coincide. Credibility concerns the audience's assessment of the communicator's honesty, independence, judgment, or reliability, whereas influence also depends on liking, identification, aspiration, attention, perceived similarity, entertainment value, and the willingness to remain exposed to the communicator over time. A highly critical influencer may be regarded as unusually discerning and therefore highly credible, while simultaneously becoming less enjoyable, less socially attractive, or less relevant to audiences who prefer affirmation, discovery, or aspirational content.

Let $T(n)$ denote credibility and $L(n)$ denote some broader measure of liking, affinity, or parasocial attraction. Overall influence can be represented schematically as

$$F(n) = T(n)^{\alpha} L(n)^{\beta}, \tag{64}$$

where $\alpha, \beta > 0$ indicate the relative importance of credibility and affinity within a particular context. If credibility follows the single-peaked form proposed earlier, while liking begins to decline at a lower or different rate as negativity increases, then the influence-maximizing point

$$n_F^* = \arg\max_n \; F(n) \tag{65}$$

need not equal the trust-maximizing point

$$n_T^* = \arg\max_n \; T(n). \tag{66}$$

Indeed, it is plausible that

$$n_F^* < n_T^* \tag{67}$$

in many influencer settings, since an audience may continue to perceive additional criticism as evidence of discernment after the same criticism has already begun to reduce affinity or enjoyment. Figure **??** illustrates one such configuration, in which affinity declines earlier than credibility, so $n_F^* < n_T^*$. The reverse ordering is also theoretically possible in contexts where criticality itself attracts attention or strengthens identification with an audience that values skepticism, which again suggests that the relative position of the two optima should be treated as an empirical question rather than an assumption.

This distinction becomes particularly important when considering the commercial interpretation of deinfluencing. An influencer who maximizes trust may not maximize purchase behavior, engagement, retention, or recommendation acceptance, because the strongest possible signal of independence may also weaken the interpersonal or aspirational qualities through which influence operates. A rare positive recommendation issued by a highly critical reviewer may possess exceptional credibility at the moment of endorsement, while the creator's overall audience may be smaller or less emotionally attached. Conversely, a more positive influencer may retain a larger or more

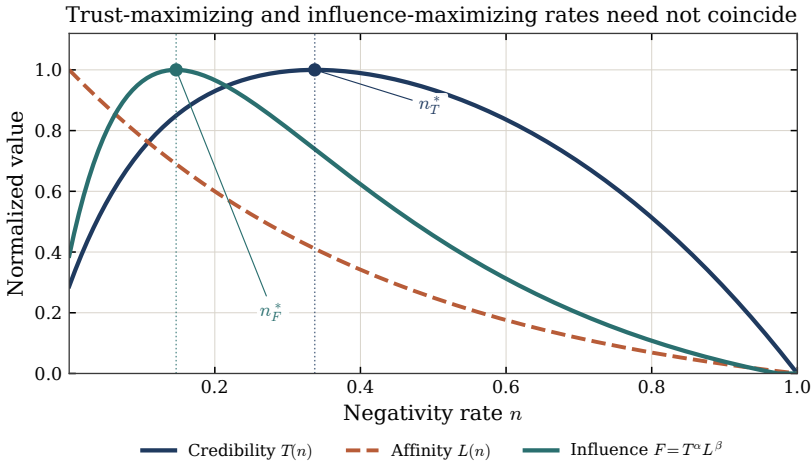

Figure 4: Schematic comparison of credibility $T(n)$, affinity $L(n)$, and overall influence $F(n) = T(n)^\alpha L(n)^\beta$. When liking falls with negativity faster than credibility does, the influence-maximizing rate lies to the left of the trust-maximizing rate. Curves are normalized for visual comparison; the functional form of $L(n)$ is illustrative.

engaged audience while making each individual endorsement less diagnostic. The commercially relevant optimum therefore depends on the outcome being maximized, and the model should resist collapsing credibility, persuasion, engagement, and long-term influence into a single dependent variable.

The distinction also provides a useful interpretation of preference alignment and confirmation bias. A negative judgment can increase perceived honesty while decreasing perceived expertise or similarity, particularly when the criticized object is strongly valued by the observer. The influence consequence of that judgment will therefore depend on the relative importance of these dimensions for subsequent behavior. In some cases, increased honesty may outweigh disagreement; in others, the loss of perceived judgment or affinity may dominate. The same act of deinfluencing can therefore move the observer upward on one evaluative dimension and downward on another, reinforcing the argument that an aggregate measure of trust may conceal important internal trade-offs.

The empirical agenda that follows from the model should therefore focus less on simple contrasts between deinfluencing and non-deinfluencing and

more on estimating response surfaces across meaningful ranges of negativity. A basic study could manipulate the proportion of negative evaluations within an influencer's visible history, using several points such as

$$n \in \{0, .10, .25, .50, .75, .90\}, \tag{68}$$

and estimate nonlinear effects on perceived independence, honesty, expertise, contrarianism, liking, purchase intention, and willingness to continue following the influencer. The primary objective would be to determine whether the expected attributional sequence emerges and whether the resulting credibility function is approximately single-peaked.

Future studies could then introduce moderator variables implied by the theoretical framework. Preference alignment can be manipulated independently of negativity frequency by varying whether the influencer's negative judgments concern products that participants already like or dislike. Identity salience can be manipulated by comparing functionally similar products with objects more strongly associated with taste, status, community, or self-definition. Perceived strategic intent could be altered by informing participants that deinfluencing is increasingly used as a creator strategy, thereby testing whether greater awareness of strategic contrarianism reduces the diagnostic value of negativity.

Longitudinal designs would be particularly valuable because the model is explicitly concerned with accumulated evaluative history. A single negative review may produce a different inference from the same review embedded in a sequence of prior judgments, and repeated exposure allows researchers to examine whether audiences update gradually, whether the effect of early criticism persists, and whether a transition from independent to contrarian attribution occurs after a sufficiently sustained pattern. Such designs would also allow researchers to distinguish immediate response from longer-term reputation formation.

Observational research could complement experimental work by estimating real-world relationships between the frequency of negative content and subsequent audience responses. Creator histories could be coded for valence, sponsorship, category, and evaluative severity, while outcomes such as engagement, follower growth, endorsement response, or audience sentiment could

be modeled nonlinearly. Such designs would face substantial selection and endogeneity problems, particularly because creators choose when to criticize and may adapt their content in response to prior audience feedback, but these complications are themselves theoretically relevant because the model predicts strategic adaptation and feedback between audience interpretation and influencer behavior.

The three-type Bayesian model also generates a more specific empirical program. Rather than measuring credibility alone, studies can ask participants to estimate the perceived probability that an influencer is primarily promotional, independently evaluative, or strategically contrarian after observing different distributions of positive and negative content. If the proposed mechanism is correct, increasing negativity should initially reduce the promotional attribution, then increase the independent attribution, and eventually increase the contrarian attribution. Estimating these posterior-like judgments directly would provide a stronger test of the theory than demonstrating an inverted-U in trust without examining the interpretive process that produces it.

Further work could also investigate whether audience members differ systematically in the parameters of the credibility curve. Individuals with high advertising skepticism may obtain more information from early negative judgments than individuals who begin with greater trust in influencers. Audiences accustomed to critical review cultures may tolerate substantially higher negativity before assigning a contrarian interpretation. Similarly, long-term followers may possess priors based on a much richer history than first-time viewers, producing different responses to identical current behavior. Hierarchical or latent-class models would be useful for estimating such heterogeneity rather than treating the observed population curve as if it represented a single shared response function.

The effects of strategic contrarianism deserve particular empirical attention because they imply that the credibility value of deinfluencing may change as the practice becomes culturally familiar. Researchers could compare audiences with different levels of awareness of deinfluencing as a marketing strategy, manipulate the perceived commercial cost of issuing a negative review, or examine whether criticism directed toward a potential sponsor is treated as more diagnostic than criticism carrying little economic consequence. These studies

would help determine whether perceived sacrifice restores signal value when audiences are already aware that negativity can be performed strategically.

The broader implication of the model extends beyond influencer marketing. The same structure may apply wherever credibility is inferred from an actor's willingness to disagree with expected incentives, including political commentary, journalism, expert review, organizational communication, professional advice, and interpersonal recommendation. Selective dissent can function as evidence of autonomy because it demonstrates that agreement is conditional, while excessive dissent can reduce informational value once disagreement itself becomes predictable. The deeper theoretical object is therefore the social interpretation of patterned agreement and disagreement.

Seen in this wider sense, the credibility problem concerns the informational value of selective opposition. Perfect agreement provides little evidence of independent judgment when agreement is expected or rewarded, while persistent disagreement also becomes weakly diagnostic if opposition appears habitual or strategically useful. Credibility is more likely to emerge when audiences believe that either agreement or disagreement could plausibly occur and that the direction of judgment depends meaningfully on the object being evaluated. Deinfluencing offers a particularly visible contemporary case of this more general process because the contrast between promotion and rejection is easy to observe and because audiences are increasingly aware of the commercial incentives surrounding influencer behavior.

References

[1] COSKUNER-BALLI, G., CHU, C., MYHR, N. and NISTOR, C., 2026. Deinfluencers as institutional actors: dramaturgical work and the embedded contestation of consumerism. *Journal of Business Research* [online]. vol. 210, 116152. Available from: https://doi.org/10.1016/j.jbusres.2026.116152

[2] ELHAJJAR, S. and ITANI, O.S., 2025. Examining the impact of social media de-influencing on audiences. *Internet Research* [online]. Available from: https://doi.org/10.1108/INTR-04-2024-0574

[3] HARFF, D. and LE, D.P.N., 2026. "Don't buy it!" Influencers' negative product reviews and their effects on young adults. *Journal of Interactive Advertising* [online]. vol. 26, no. 2, pp. 96–103. Available from: https://doi.org/10.1080/15252019.2026.2615713

[4] HUSSAIN, A., 2026. The rise of de-influencing: authenticity, consumer resistance, and the governance of digital persuasion. *EDPACS* [online]. Available from: https://doi.org/10.1080/07366981.2026.2681070

[5] KHALIL, T., MAHMOOD, C.K., SINGH, H., AFTAB, F., KHAN, H., MANZOOR, I. and ABDULHAYEE, K.A., 2025. The digital de-influencing wave: redefining trust in online communities. *Frontiers in Communication* [online]. vol. 10, 1600657. Available from: https://doi.org/10.3389/fcomm.2025.1600657

[6] KOIVUNEN, K., HAANPÄÄ, M.A.A. and SARANIEMI, S., 2025. The emergence of cause-oriented influencers – conceptualizing de-influencing on TikTok. *Journal of Business Research* [online]. vol. 197, 115463. Available from: https://doi.org/10.1016/j.jbusres.2025.115463

[7] MICHAELIDOU, N., KOSTOPOULOS, I. and LOWE, E., 2026. The dawn of 'deinfluencing' as a vehicle for moral responsibility and anti-consumption. *Journal of Business Research* [online]. vol. 202, 115773. Available from: https://doi.org/10.1016/j.jbusres.2025.115773

[8] PENTTINEN, V., CIUCHITA, R. and PEDERSEN, S., 2026. Let me deinfluence you! The impacts of deinfluencing content. *Journal of the Academy of Marketing Science* [online]. Available from: https://doi.org/10.1007/s11747-026-01141-9

[9] TRAN, G.A., ABOULNASR, K. and SONG, Y.S., 2025. The rise of the deinfluencers: perceived similarity and anticipated emotions in advocating mindful consumption on social media. *Psychology & Marketing* [online]. vol. 42, no. 11, pp. 3005–3026. Available from: https://doi.org/10.1002/mar.70024

About the Author

Robin Dub is a professor of marketing in Florida. His work focuses on intercultural communication and global markets. Of Czech descent, he also turns to the most varied areas of life in Central Europe, past and present. His research includes consumer behaviour and digital marketing.

Robin Dub (Miami)
ORCID 0009-0006-4583-4764