Scientific and Social Research, 2026, Volume 8, Issue 4
http://ojs.bbwpublisher.com/index.php/ssr
Online ISSN: 2981-9946
Print ISSN: 2661-4332

Why Debunking Fades: Misinformation Correction as a Temporary Achievement in Competitive Social Media Environments

Qinyu E, Xiaoying Huang

College of Publishing, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Research on misinformation correction has focused mainly on whether debunking reduces false beliefs immediately after exposure, while paying less attention to why such effects weaken over time. This paper argues that correction effectiveness is best understood as a temporary persuasive achievement whose durability depends on continued success in post-exposure competition. Corrective influence must remain competitive across three linked domains: attention allocation, interpretive authority, and memory retrievability. Corrections fade when they lose attention, lose control over how they are understood, and become less retrievable than the misinformation they seek to displace. Debunking decay, in this sense, reflects the difficulty of sustaining corrective influence over time.

Keywords: Misinformation correction; Social media; Debunking decay; Belief regression

Online publication: May 25, 2026

1. Introduction

The rapid spread of misinformation on social media has made debunking a central concern in communication research. Existing studies have mainly asked whether corrective messages reduce misperceptions or improve belief accuracy at the moment of exposure. Although this work shows that corrections can produce meaningful short-term gains, those gains are often less durable than immediate post-exposure measures suggest [1–2].

A growing body of research indicates that corrective effects do not necessarily remain stable over time. Under some conditions, belief improvements weaken, corrected claims continue to shape later judgment, and the original misinformation remains cognitively available even after it has been rejected [2–3]. The key question, therefore, is not only whether correction works, but whether its influence can endure over time.

This paper argues that correction effectiveness is best understood as a temporary persuasive achievement. On social media, corrections circulate in environments shaped by information overload, rapid turnover, social commentary, and repeated exposure to competing claims. Correction is therefore not

complete once belief shifts; it must continue to compete with rival messages, meanings, and memory traces.

To explain this process, the paper conceptualizes debunking decay as a competitive post-exposure dynamic. More specifically, corrective influence depends on continued success across three sequential but interacting competitions. First, corrections must attract and hold attention in environments saturated with vivid, affectively engaging, and algorithmically prioritized content [4]. Second, they must secure interpretive authority in socially contested spaces, where corrective meaning can be weakened both by the persistence of negative impressions after factual updating and by surrounding commentary, stance cues, and social reactions that reshape how the correction is understood [1–2]. Third, corrections must remain cognitively retrievable over time, so that the correction rather than the original false claim is available when individuals later make judgments [3, 5]. When corrections lose ground in these competitions, what initially appears to be successful debunking may gradually fade.

2. From immediate updating to corrective durability

To theorize debunking decay more precisely, it is necessary to distinguish among several related but non-identical forms of post-correction persistence. The continued influence effect provides the broadest theoretical backdrop, referring to the tendency for misinformation to continue shaping reasoning and judgment even after correction.[3] Within this broader problem, belief regression refers to the temporal weakening of corrected belief: individuals revise their judgments after correction, but later drift back toward the original false claim [3]. Belief echoes, by contrast, refer to the persistence of attitudinal or affective residue even after factual belief has been updated [1].

These distinctions matter because corrective incompleteness does not take a single form. Misinformation may remain influential because individuals return to the false claim itself, because evaluative residue survives factual updating, or because corrected information continues to shape later reasoning in more diffuse ways. Without distinguishing these possibilities, it becomes difficult to specify what exactly is fading after correction and why.

Among these forms of post-correction persistence, this paper focuses on corrective durability: whether an initially successful correction remains influential over time. In this framework, belief regression is the most direct indicator that corrective durability has weakened, because it captures the temporal loss of corrected belief [3]. Belief echoes, in turn, are treated not as an alternative outcome of equal status, but as one important mechanism through which durability may be undermined, especially at the level of interpretation and evaluation [1]. The continued influence effect serves as the broader context for understanding why correction may remain incomplete even after initial updating.

This distinction is especially important in social media environments, where correction does not end at the moment of exposure. Users do not encounter misinformation once, revise their beliefs, and then leave the issue behind. They continue to move through information streams filled with new content, alternative framings, social reactions, and repeated reminders of the original claim. The post-exposure environment thus becomes a decisive stage: it is where an initially successful correction may either stabilize into a durable influence or gradually weaken. From this perspective, the decay of debunking effectiveness can be defined as the process through which corrective influence weakens after initial exposure because it can no longer remain competitive in the post-exposure environment. Debunking decay, in this sense, is not a single point of failure, but the gradual erosion of corrective influence after initial success.

3. Competition one: Attention allocation

The first competition takes place at the level of attention. If corrective influence is to endure beyond the moment of exposure, it is not enough for a correction to be merely noticed. It must attract sufficient attention to be processed, integrated, and encoded strongly enough to matter later. This does not mean that belief change cannot occur under limited attention. A correction may still prompt initial updating. But when attention is weak or brief, that change is more likely to remain shallow and fragile, making it easier to erode over time.

This first competition is difficult for two reasons. First, corrections often enter the attention race under less favorable conditions than the misinformation they challenge. Second, even when they are noticed, they usually require more effortful processing, which makes durable encoding harder to achieve in fast-moving social media environments.

3.1. Fluency asymmetries

One reason corrections struggle for attention is that the competition is often asymmetrical from the outset. Misinformation may appear in forms that are more vivid, immersive, or easy to process than the correction that follows it [4]. The issue is not that any particular modality is inherently misleading, nor that corrections must always be text-based. The issue is comparative: when misinformation is presented in a format that is more salient, rewarding, or fluently processed than the correction, the two messages do not compete on equal terms.

Importantly, this asymmetry does not make correction impossible. Corrections can still produce belief change, especially when they are timely, credible, and clearly presented. What asymmetry does is place correction at an attentional disadvantage from the start. As a result, even when a correction shifts belief initially, that shift may rest on a weaker processing foundation than the misinformation it seeks to counter. The problem, then, is not that correction cannot work, but that it often begins from conditions that make its effects less stable.

This matters especially on platforms built around fast scrolling, continuous novelty, and fragmented exposure. In such settings, users often allocate attention according to salience, affective pull, and processing fluency rather than epistemic value alone. A correction may therefore be accurate and well supported yet still struggle to hold attention long enough to establish a durable foothold.

3.2. Cognitive effort and shallow processing

Corrections also tend to demand more cognitive effort than the misinformation they challenge. Processing a correction often requires users to do more than register a new message. They may need to recognize a contradiction between the original claim and the corrective response, assess the credibility of the correction, and revise an existing understanding of the issue in light of new evidence. In this sense, correction requires evaluative processing rather than mere exposure.

This becomes especially difficult in social media environments, where users are often cognitively busy. They encounter corrections while multitasking, moving quickly through a feed, or engaging with content designed for entertainment rather than reflection [4]. Under these conditions, even noticed corrections may receive only shallow processing. Users may register that a claim has been disputed without fully encoding why it is false or how the correction should replace the original belief.

The consequence is weak initial encoding. Because later durability depends partly on the depth of early processing, a correction that is only shallowly processed is less likely to remain retrievable, resist later social reframing, or survive repeated reminders of the original misinformation. For this reason, decay begins not only when a correction is later forgotten, but earlier, when it fails to receive the depth of attention needed to support durable influence.

4. Competition two: Interpretive authority

Winning attention is not enough for a correction to remain effective. After a correction is noticed, a further question arises: what does it come to mean for those who encounter it? On social media, users rarely encounter a correction in isolation. They see it together with comments, reposts, stance cues, and evaluative reactions. These surrounding responses matter because they shape not only whether the correction is noticed, but how it is understood. Debunking, therefore, competes not only for visibility but also for interpretive authority. A correction begins to lose influence when it is still present as information but no longer serves as the main basis through which people make sense of the issue.

4.1. Belief echoes and residual evaluation

One reason interpretive authority matters is that correcting a falsehood does not necessarily erase the impression it leaves behind. This is the logic of belief echoes. A correction may change what people regard as factually true while failing to remove the broader evaluative meaning generated by the original misinformation. In such cases, the false claim no longer survives as an accepted fact, but it continues to affect how a person, event, or institution is judged.

A simple example makes this clearer. Suppose a rumor claims that a political candidate accepted a bribe. A later fact-check shows that the accusation is false. Many people may then accept that the specific claim is untrue. Yet some may still come away with the sense that the candidate seems suspicious, untrustworthy, or somehow compromised. The allegation is no longer believed as fact, but its negative evaluative residue remains.

This helps explain why a correction can weaken even when it is not directly rejected. The problem is not always that users deny the correction or continue to endorse the false claim. Sometimes the correction succeeds at the level of factual belief but fails at the level of broader judgment. It changes what people think happened, but not fully how they evaluate the target of the misinformation. When that happens, interpretive competition has not truly been won. The correction has updated the fact, but it has not fully displaced the meaning attached to the original claim.

4.2. Social cues and contested meaning

Interpretive authority is also shaped by the social context in which a correction is encountered. Users rarely engage with a fact-check in a neutral environment. Instead, they see it alongside peer responses, visible stance cues, partisan framing, and commentary that suggests how the correction should be interpreted. These surrounding reactions do not merely accompany the correction. They actively participate in defining what the correction means and whether it should be trusted.

For example, a correction may clearly state that a viral claim is false and that the supporting evidence has been manipulated. But if the surrounding comments say things like “this is just damage control”, “the

denial makes it even more suspicious", or "everyone knows what is really going on", many users will not encounter the correction as a straightforward factual clarification. They will encounter it as a socially reframed message. In this way, comments and reactions do more than add noise. They help tell users what the correction stands for and how seriously it should be taken.

Once interpretive authority shifts away from the corrector, the influence of the correction weakens. People no longer rely primarily on the correction itself to understand the issue. Instead, they rely more on peer reactions, identity-congruent cues, or socially dominant frames. The correction remains visible, but its meaning becomes contested. When that happens, its capacity to guide later judgment is reduced, even if the correction itself is still remembered.

The key point, then, is that interpretation is not secured once a correction is seen. It must remain socially defensible after exposure. In this sense, debunking decay reflects not only the fading of information but also the gradual erosion of interpretive authority in environments where meaning is continuously negotiated.

5. Competition three: Memory retrievability

The third competition is about a simple but important question: when people return to the issue later, do they still remember the correction? A correction may work at the moment of exposure, but it can still lose its effect over time if people no longer think of it when the original claim comes up again. In this sense, the problem is not only whether the correction was accepted at first, but whether it still comes to mind when people make later judgments. Corrective influence weakens when the correction becomes harder to recall than the misinformation it was meant to replace.

5.1. Belief regression and memory failure

This matters because belief regression is often closely related to memory failure. People may change their minds immediately after seeing a correction, but later drift back toward the original false claim. This does not always mean that they have deliberately rejected the correction. Sometimes the simpler explanation is that, by the time they encounter the issue again, they no longer clearly remember the correction.

Research by Swire-Thompson et al. (2023) makes this point especially clear. Their findings show that memory for whether a claim had been corrected explains a substantial share of the variation in belief regression [3]. In other words, later reversion often happens not because people actively decide to believe the misinformation again, but because the correction is no longer strong enough in memory to guide their judgment.

This changes how people should understand decay. If correction weakens mainly because memory weakens, then later regression should not always be treated as resistance or motivated rejection. In many cases, people may move back toward misinformation simply because the correction no longer comes to mind when they need it.

5.2. Unequal recall in social media environments

This problem becomes more serious on social media because the original misinformation is often repeated, while the correction is not. People may see the false claim again in reposts, screenshots, jokes, memes, or casual discussions. Each reappearance makes the original claim easier to remember. By contrast, the correction is often encountered only once and then left behind. As a result, the original misinformation may

remain easier to recall than the correction. This does not mean that people fully believe the rumor again in a deliberate way. Rather, when they think about the issue later, the false claim is simply more likely to come to mind first. Once that happens, later judgment is more likely to lean back toward the misinformation.

The key point, then, is that successful correction requires more than initial acceptance. It also requires that the correction remain easy enough to remember when the issue comes up again. In this sense, debunking decay reflects not only fading belief change over time, but also the growing difficulty of bringing the correction back to mind when it is needed.

6. Conclusion: A sequential and interacting process model

The preceding discussion shows that corrective influence may weaken at three closely related points: when a correction fails to attract sufficient attention, loses authority as a guide to interpretation, or becomes less retrievable than the misinformation it seeks to displace. These are not isolated weaknesses, but part of a broader process through which initially successful debunking gradually loses durability.

Correction decay is therefore best understood as a sequential but interacting process rather than a simple accumulation of separate factors. It is sequential because corrective influence must first attract enough attention to be encoded, then retain enough interpretive authority to guide understanding, and finally remain retrievable enough to shape later judgment. It is interacting because weakness at one stage can intensify vulnerability at the next, while later developments can reopen earlier competitions.

The process often begins with weakness at the level of attention. When a correction is only briefly noticed or shallowly processed, its initial encoding is fragile. Immediate belief change may still occur, but it is less likely to endure when users later encounter competing content, social reactions, or renewed exposure to the original misinformation. This fragility becomes more consequential at the interpretive stage, where a correction may remain visible and even remembered, yet still lose influence if it no longer shapes how the issue is understood. Interpretive erosion does not simply follow attentional weakness; it can also amplify it by making the correction seem less relevant, less credible, or less worth retaining. Over time, especially under repeated exposure to the original claim, the correction may become less likely to enter later judgment. At that point, the problem is not that the correction never worked, but that it is no longer cognitively competitive. Once misinformation becomes easier to retrieve than the correction, belief is more likely to drift back toward the false claim.

These stages are not fixed or strictly one-directional. Repeated exposure to misinformation can reopen interpretive competition by inviting renewed commentary, mockery, or partisan reframing. In turn, interpretive erosion can accelerate memory decline by reducing the likelihood that users will continue to rely on or rehearse the correction. Debunking decay is therefore best seen as a dynamic trajectory in which corrective influence gradually loses ground across linked competitions in the post-exposure environment.

Research on misinformation correction shows that false beliefs can be revised in the short term, but immediate belief change does not guarantee durable correction. The challenge of debunking, then, is not only to correct misinformation once, but to make that correction durable enough to withstand the continuing pressures of social media environments. Future research should therefore focus not only on whether correction works at one moment, but also on the conditions under which it can endure over time.

Funding

This work was supported by the National Social Science Fund of China (Youth Project) under Grant No. 25CXW016.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Chae JH, Groeling T, Song H, 2026, Time is the Fire in which Message Effects Burn: Decay and Sustenance of Correction Effects over Time. Journal of Communication, 76(1): 36–48. https://doi.org/10.1093/joc/jqaf030

[2] Nyhan B, 2021, Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions. Proceedings of the National Academy of Sciences, 118(15): e1912440117. https://doi.org/10.1073/pnas.1912440117

[3] Swire-Thompson B, Dobbs M, Thomas A, & et al., 2023, Memory Failure Predicts Belief Regression after the Correction of Misinformation. Cognition, 2023(230): 105276. https://doi.org/10.1016/j.cognition.2022.105276

[4] Gunasekara S, Sew C, Pareek S, et al., 2026, Timing Matters: Designing Effective Corrections for Short-form Video Misinformation. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26). ACM. https://doi.org/10.1145/3772318.3790560

[5] Capewell G, Maertens R, Remshard M, et al., 2024, Misinformation Interventions Decay Rapidly without an Immediate Posttest. Journal of Applied Social Psychology, 54(8): 441–454. https://doi.org/10.1111/jasp.13049