⟡ ChatGPT

Stress-Testing the "Eight Blockers" Thesis of Belief Resistance

Introduction

Why do facts and logic often fail to change the minds of people immersed in online misinformation? The thesis under review claims there are **eight fundamental internal "blockers"** – cognitive or affective mechanisms inside the person – that explain why disinformation believers resist corrective evidence. These range from limited cognitive bandwidth to social identity pressures. Here we critically evaluate each of the eight proposed blockers against current research and consider whether any additional internal factors might be missing. The focus is on **internal** psychological processes (not external platform factors), and each blocker should be distinct, detectable, causally impactful, and targetable by specific interventions. We draw on recent reputable studies in psychology, communication, and political science to verify the validity and independence of these blockers. Overall, the eight-factor framework aligns well with known drivers of misinformation persistence [1] [2] , though we discuss nuances, overlaps, and potential omissions (such as overconfidence or reactance).

1. Bandwidth and Cognitive Capacity

Thesis definition: Limited attention, working memory, or high ambiguity leads people to take mental shortcuts or avoid effortful belief updates. In other words, when processing capacity is low (due to fatigue, overload, complexity) or information is unclear, individuals will be less likely to absorb logical corrections. This blocker is said to operate even absent identity or emotional factors.

Evidence: There is strong support that reduced cognitive bandwidth or effort can undermine fact-processing. Research shows that people who rely on intuition and don't engage in analytic thinking are more prone to believing false news, *even when partisan bias would predict otherwise*. In fact, susceptibility to fake news is "better explained by lack of reasoning than by motivated reasoning" in many cases [3] . Conversely, prompting more deliberation and attention improves truth discernment: experiments found that inducing people to slow down and think (engage analytical capacity) significantly reduces belief in false headlines [4] . Cognitive load studies echo this – being mentally busy or distracted increases the chances of accepting misinformation. For example, a recent experiment using eye-tracking noted that heavy **cognitive load made participants more likely to believe fake news** and impaired their ability to distinguish it from real news [5] . All of this suggests that a **"bandwidth" constraint is a real internal blocker**: when our brains are tired or overloaded, we default to shallow processing and are less receptive to corrective logic. This effect does not require any social identity or emotions; it can occur even for neutral topics, which supports the thesis that it's non-reducible to the other factors.

Distinctness: Bandwidth limitations are distinct from other blockers in that they are about *mental capacity and effort*, not what the information *means* to the person. Even a person with no group loyalties or emotional stake can fail to update a belief if the correction is too complex or if they're mentally fatigued. One isolation test of this idea is varying the clarity and cognitive load of a message while holding content

constant – indeed, acceptance of the message tracks the cognitive load, as the thesis predicts. This blocker also connects to the **"lazy not biased"** finding: many misinformation believers aren't deeply convinced ideologues so much as cognitively misers [3] . When information is presented in simpler, more digestible formats (e.g. clear summaries, infographics), their acceptance of facts tends to increase, which is a "tell" that bandwidth was a limiting factor. In sum, **research validates limited bandwidth as a fundamental blocker** that operates alongside (and independently from) social or motivational factors.

2. Belonging (Social Alignment)

Thesis definition: A drive to remain aligned with one's "in-group" can override factual persuasion. Even without any immediate peer pressure, people have an internalized bias toward beliefs that signal loyalty to their community or tribe. *Tells:* The same evidence is accepted when it comes from an in-group member but resisted when coming from an out-group or rival source. This blocker is distinguished from public reputation (it operates even in private) and from personal identity content (it's about *who* is providing the information).

Evidence: Decades of social psychology research confirm that source group identity strongly influences persuasion. **Individuals are typically more persuaded by messages from an in-group source than by identical messages from an out-group source** [6] . For example, Natalie Wyer and colleagues showed that when people received an argument, they were swayed more if they thought it came from someone sharing their political or social group, versus from an outsider [6] [7] . This holds especially true when the group identity is salient and relevant to the issue. Another classic demonstration: partisans will flip their policy opinions depending on which political party is said to support the policy – even if the policy content is unchanged. In misinformation contexts, one study found that **corrections paired with in-group endorsers (such as a trusted community leader or a co-partisan expert) significantly increased acceptance**, whereas the same information from an out-group figure was met with skepticism or indifference [6] . These findings align perfectly with the belonging blocker: people have an internal social *homing signal* for what "our people" believe, and they resist being the odd one out.

Notably, this effect persists in private settings. Even when no other group members are watching, individuals may still reject a fact that, if true, would put them at odds with their group's consensus. This suggests an internalized social identity motive (often called **"identity-protective cognition"** in a group sense) rather than mere fear of immediate social sanction. The belonging mechanism is **not reducible to reputation management** – studies have documented in-group bias in information processing under anonymous conditions, indicating it's about psychological attachment, not just public image [6] . It's also distinct from personal ideology (next blocker) because it's specifically tied to *who is saying it*. In short, **the "Belonging" blocker has robust empirical support**. It highlights the importance of *messenger identity*: the exact same factual correction can fail or succeed based solely on whether the audience perceives the source as "one of us" or "one of them." This remains one of the most potent reasons why logical arguments (from the "wrong" messenger) bounce off certain audiences.

3. Reputation Preservation (Audience Costs)

Thesis definition: This blocker refers to the reluctance to admit error or change one's stance due to concern for one's social reputation and consistency in the eyes of others. If a person has publicly committed to a false belief, acknowledging the truth can risk embarrassment, loss of face, or status – so they double

down or remain silent rather than concede. *Tells:* We might see a person give different answers in public versus in a private message, or observe that as soon as one respected peer voices doubt, others quickly follow (because the reputation cost of agreement drops).

Evidence: The power of public commitment and consistency in shaping attitudes is well-documented in social psychology. **Once people make a stance public, they become significantly more resistant to later persuasion that contradicts their stated position** [8] **.** This is rooted in our desire to appear consistent and avoid the negative judgements that come with flip-flopping. Research shows that *public commitments "tie our reputation" to that position*, thus creating an internal pressure to stick with it [8] . One review put it succinctly: **public declarations engage impression management motives – after taking a stand, individuals strive to appear consistent to maintain a favorable image and avoid looking hypocritical** [9] [8] **.** In practice, this means even if someone privately recognizes they might be wrong, they may continue defending the false belief if an about-face would embarrass them in front of their peers or followers.

For example, consider social media behavior: Sacha Altay and colleagues found that **people share relatively little fake news on their personal feeds because doing so risks their reputation; they would share more only if incentivized or anonymized** [10] **.** This implies many know at some level that certain information is dubious, but *openly correcting or retracting it* would mean admitting a mistake, which they avoid. In group discussions, the first person to break from the group line often gives cover for others to agree – highlighting that reputational fear was suppressing honest belief revision until it became acceptable. The thesis's isolation test (public vs private responses) is echoed by findings that survey answers on sensitive topics differ when anonymous versus identified, showing people will privately acknowledge facts that they won't concede publicly due to social repercussions.

Distinctness: Reputation concerns are related to but separable from pure group loyalty. You might privately disagree with your group's misinformation, yet publicly nod along to avoid social penalty – that's reputation preservation at work. It's also distinct from internal identity belief; here the person might intellectually know the facts, but the blocker is the *social cost* of saying so. The commitment-consistency principle underscores this: once committed, **people often experience cognitive dissonance at the idea of reversing themselves, and they resolve it by rejecting new information that implies they were wrong** [8] [11] . In summary, research strongly supports that **fear of humiliation or status loss is a key internal barrier** to changing one's stated beliefs. Neutralizing this (say, by allowing face-saving off-ramps or private belief change) can therefore make factual updates more palatable.

4. Identity Preservation (Protective Cognition)

Thesis definition: This blocker is the classic *"motivated reasoning"* phenomenon: people resist facts that clash with their core identity, values, or worldview. If accepting a piece of information means acknowledging that "my ideology/faith/values were wrong" or feeling like a traitor to one's prior self-concept, the mind will generate counter-arguments and rationalizations to protect the identity. Unlike belonging, which is about group allegiance, identity preservation is about *self-image and deeply held worldviews*. It operates even in private, with no audience – it's an internal defense mechanism against cognitive dissonance and identity threat.

Evidence: The literature overwhelmingly supports identity-protective cognition as a driver of factual disbelief. Dan Kahan, who coined the term, summarizes it: **individuals tend to "selectively credit and**

dismiss evidence" based on whether it affirms or threatens their defining group values or personal worldview [12] **.** In other words, people unconsciously *reject information that implies their prior beliefs or group's position is wrong*. A concrete example comes from experiments on politically polarizing issues: when a scientific finding or news report contradicted participants' ideological stance, they often found ways to disbelieve or downplay it, yet would accept similarly strong evidence when it aligned with their identity. **Kahan notes that people are more likely to accept misinformation and resist corrections** *when the misinformation is identity-affirming rather than identity-threatening* [13] **.** This aligns with the thesis claim: a false belief that "fits" your worldview feels emotionally safer to hold, and you'll find ways to discount even solid logic that undermines it.

Numerous studies on issues like vaccine denial, climate change, or conspiracy theories find that identity often trumps factual accuracy. For instance, a study on correcting political misperceptions found that corrections often fail or even backfire among those for whom the false belief is tied to their political identity [14] . People engage in motivated skepticism: scrutinizing disconfirming evidence harshly but uncritically accepting information that supports their pre-existing view. Neuroimaging research has even shown that challenges to political beliefs activate brain regions associated with negative emotion and pain, reflecting how *threatening it feels to the self* to have cherished beliefs challenged. This blocker is essentially an internal *emotional* filter: logic that implies "your team/ideology is wrong" triggers defensive reasoning rather than open reception.

Distinctness: Identity preservation is closely related to the belonging blocker but can be disentangled. One can imagine a scenario with no salient in-group/out-group messenger (say, reading an anonymous scientific report alone) – if the findings conflict with your worldview ("implies who I am or what I've believed is wrong"), identity-driven denial can still kick in. The thesis gives an isolation test: low audience and high clarity, vary whether the correction attacks identity. Indeed, experiments with *refutation texts* show that when the topic is identity-charged (e.g., it implicates religion, ideology, or self-concept), learners show far less belief revision, even under optimal learning conditions [15] . When the same style of correction is applied to a topic that isn't identity-relevant, people update more easily. This confirms identity threat as a unique blocker. It's not reducible to reputation (since it occurs even with no witnesses) and not exactly the same as group belonging (since it's about personal values as much as group loyalty). **In sum, identity-protective cognition is a foundational reason why facts fail to penetrate – well-supported by research on motivated reasoning and cognitive dissonance.** If an intervention can remove the sting of identity threat (for example, by affirming people's values *before* giving a correction, or framing the truth in a way compatible with their identity), acceptance of facts often improves.

5. Authority Cues (Obedience and Credibility Heuristics)

Thesis definition: Symbols or signals of authority – titles, credentials, "official" formatting, verification badges, confident tones – can short-circuit scrutiny and lead people to accept claims without logical evaluation. Essentially, humans have a heuristic to defer to perceived authority or expertise. In the context of misinformation, this means that *if false information comes wrapped in the trappings of authority, people are more likely to believe it*, and conversely, true information from a low-status or distrusted source may be ignored. This blocker is internal in that it's a built-in cognitive shortcut (a peripheral route to persuasion). It's distinct from identity because it's about *formal or role-based trust* (e.g. doctors, officials) rather than group identity per se.

Evidence: Persuasion research strongly supports the impact of authority and source credibility cues on belief. **Source credibility has been found to influence persuasion through multiple mechanisms** [16] **:** it can bias how people process the content (we scrutinize less if the source is "expert") and also serves as a simple heuristic ("if *Dr. XYZ* said it, it must be true"). Classic experiments by Hovland in the 1950s showed that identical information was far more convincing when attributed to a high-credibility source (e.g., a prestigious scientist or institution) than to a low-credibility source. More recent studies continue to demonstrate this. For example, a 2019 study on misinformation corrections found that **when a retraction came from a source high in expertise/trustworthiness, people were much more likely to accept the correction and drop the false belief, compared to when the retraction came from a dubious source** [17] **.** Another experiment varied whether health misinformation on social media was corrected by an "expert" source (a doctor or medical organization) versus a generic source – the expert-labeled corrections were significantly more effective in changing attitudes toward vaccines [18] .

On the flip side, purveyors of disinformation often *mimic* authority to lend false claims undue credibility. We see this when pseudoscience articles cite "studies" in official-sounding journals, or when conspiracy theorists call themselves "Dr." or use government-like logos. Many people will grant the information a presumption of truth because our brains are trained to defer to authority symbols. This is essentially the obedience heuristic that Stanley Milgram famously demonstrated (people obey figures in lab coats). In belief terms, **people often accept assertions from an authority figure or source with an official marker, even if the content is shaky – the authority cue bypasses their critical filters** [16] .

Distinctness: Authority influence is conceptually distinct from group belonging. You might trust an authority *outside* your in-group if they have credentials you respect (though often we grant authority status more to in-group experts). It's also not the same as familiarity (repetition) – it can make a *first-time claim* seem credible if an expert voice says it. The thesis asserts this blocker produces acceptance shifts without changes to identity or belonging, and research agrees: for instance, an experiment kept the message identical and from the same group, but simply added or removed an "expert endorsement" badge – the presence of the badge increased belief in the message's truth [16] . That demonstrates an independent effect of authority cues. In practical terms, **this means that one way to overcome resistance is to deliver facts via highly credible messengers** (or at least to remove the false aura of authority around misinformation). Conversely, it's a warning that even logical people can be swayed by *who* said something rather than *what* was said – an internal bias that manipulators exploit.

6. Moralized Contempt (Out-group Hate and Sacred Values)

Thesis definition: This blocker involves the target's moral-emotional orientation. When an issue or belief is "moralized" – tied into a sense of absolute right vs wrong – and the believers view those on the other side with contempt or hatred, they gain a sort of **psychic reward for rejecting contrary evidence**. The false belief becomes a badge of moral virtue or righteous anger. In this state, disconfirming facts aren't just incorrect; they are *evil* or sacrilegious. Key signs include dehumanizing or demonizing language about out-groups and a framing of the debate in terms of good vs evil. Even neutral evidence may be repelled because accepting it feels like siding with immoral others.

Evidence: Strong moral convictions do indeed alter how people process information – typically in a way that increases rigidity and intolerance. Research by Linda Skitka and others on **moral conviction** finds that attitudes held with moral fervor are more resistant to change than equally strong but non-moral attitudes [19] . One reason is that moralized beliefs engage emotions like anger, disgust, and even hatred toward

those who disagree. **For example, studies have shown that when people's attitudes on an issue are deeply moralized, they exhibit *high levels of negative emotion and disdain toward opponents*, far beyond a normal dislike** [20] **.** Moral conviction is associated with seeing the opposition as not just incorrect, but bad or corrupt. Pretus et al. (2023) even found that moralized beliefs correlate with neural and self-report indicators of *hatred* toward the opposing side – a qualitatively different state than mere disagreement [20] .

This moralized contempt can create a firewall against factual correction. If a trusted source tries to correct the false belief, but that correction threatens a "sacred value" or seems to come from the despised out-group, the person's response is moral outrage rather than consideration of the facts. They might double down, thinking accepting the fact would be a betrayal of what's right. We see this dynamic in extremist or highly polarized communities: *grievance and virtue-laden rhetoric* (e.g. "defending our values against those monsters") boosts engagement and solidifies in-group consensus [21] , but it also **"heightens engagement while repelling counter-evidence," as the thesis suggests**.

Notably, moral conviction is linked to a kind of *psychological closure*: one feels absolutely certain and justified, much less open to nuanced debate [22] . A morally convinced individual often perceives their beliefs as universal truths and is willing to accept big sacrifices or dismiss counter-arguments to uphold them [23] [20] . This absolutism means evidence to the contrary is either ignored or actively attacked as morally offensive. For instance, climate change discourse sometimes gets moralized (on both sides), with each side viewing the other as not just wrong but *immoral* – which can impede acceptance of scientific facts by those who see them as tainted by the morally bad side.

Distinctness: Moralized contempt overlaps with identity and belonging (since moral values often bind groups), but it emphasizes the *affective arousal and value-based* rejection of information. One could imagine a scenario where someone's political identity is at stake (identity blocker) versus one where they feel a deep moral outrage (moral blocker) – in practice these co-occur, but they are analytically separable. The thesis differentiates this from pure identity by noting it operates via emotional reward/punishment: the person *enjoys* rejecting the out-group's claims because of contempt. Experiments support some independence: if you hold messenger and content constant but vary the moral-emotional framing (neutral tone vs indignation and sanctimony), you can see differences in how the message is received [19] [20] . A contempt-laden narrative can polarize attitudes and make listeners less receptive to any concessions, even if identity cues are the same.

In summary, moralized affect is a powerful blocker: once an issue is in the realm of tribal morality, facts become secondary to moral signaling. The literature on *affective polarization* in politics (i.e., strong partisan hatred) reinforces this – higher affective polarization correlates with more biased reasoning and dismissal of news from the other side. Neutralizing this blocker might involve depolarizing the issue or finding ways to frame corrections in values that the target doesn't see as evil (a strategy known as moral reframing).

7. Fluency and Familiarity Bias (Illusory Truth Effect)

Thesis definition: Repetition breeds belief. This blocker refers to the cognitive bias where a statement feels more true simply because one has heard it before. Every repeated exposure increases processing fluency (it's easier to think about), and our brains misattribute that ease to truthfulness. As a result, even a logically weak or false claim can gain acceptance if it's frequently repeated in one's information environment. *Tells:* People's belief in a claim often correlates with how many times they've encountered it, rather than the

strength of evidence. Also, corrections that use a consistent phrasing may work better than varied phrasing, because familiarity helps acceptance.

Evidence: The **illusory truth effect** is a well-established phenomenon in cognitive psychology. **Simply repeating a false statement can increase the number of people who agree with it, often regardless of the source or the person's prior knowledge** [24] [25] **.** For instance, a meta-analysis by Dechêne et al. (2010) found that across many studies, repetition reliably raised perceived truth ratings of statements – this held true even for statements that were initially recognized as false [24] . In a misinformation context, Pennycook et al. (2018) demonstrated that **after just one prior exposure to a fake news headline, participants rated it as more accurate on a later encounter compared to headlines they had never seen** [25] **.** Strikingly, this effect appeared **"equally for implausible statements as for plausible ones"**, indicating how automatic and content-insensitive the familiarity bias is [26] . Even when people know better (e.g., the false statement contradicts what they firmly believe or could verify), repetition can create a lingering sense of "maybe there's something to it."

There's also evidence that processing fluency – how easy a message is to digest – contributes. Clear, smoothly worded misinformation tends to be believed more than a complicatedly phrased truth. Fluency from repetition is one driver of this ease. Unkelbach and colleagues have theorized that our brains use a heuristic: "if I recall hearing this, it might be true" (referential validity). This is why propaganda and viral misinfo place such emphasis on repeating slogans and claims. Furthermore, counter-misinformation experts warn about the "continued influence effect": once a falsehood lodges in memory, even after a correction people may continue to rely on the false detail, partly because it's more familiar than the correction if the correction isn't repeated enough [27] . Consistently phrasing the correction the same way each time can help build its familiarity and thus its credibility, aligning with the thesis that stable phrasing outperforms variable phrasing.

Distinctness: Familiarity bias is a cognitive quirk distinct from motivational blockers. It can sway even those with no stake in the claim. For example, people will believe random trivia facts or urban myths more if they've heard them before, simply due to exposure [24] . This works independently of who said it or whether it aligns with identity – it's an unconscious perception of truthiness from repetition. Notably, the illusory truth effect operates even when sources are not highly trusted, which differentiates it from authority cues; even nonsense repeated by a known unreliable source can start to feel true if heard enough (though source credibility and repetition can combine for a stronger effect) [28] .

In practical terms, this blocker means that **disinformation often succeeds by sheer volume and redundancy**. It also means debunking efforts must be careful: merely repeating the myth to refute it can inadvertently reinforce the myth's familiarity (hence strategies like "lead with the truth, then mention the myth" are recommended in debunking). The existence of fluency bias underscores that a person might intellectually know the logical arguments, but still have a nagging "gut" feeling that the false claim is true because "I keep hearing it everywhere." This is absolutely an internal blocker to fully accepting the correction, and neutralizing it requires time and repetition of the correct information (or at least exposing the person to the correction as frequently as they were to the myth). The research consensus backs the thesis claim that **familiarity breeds belief** [25] – making it a fundamental cognitive barrier to belief updating in the misinformation age.

8. Epistemic Closure (Alternate Reality Bubbles)

Thesis definition: This final blocker describes a state where the person's entire framework of trusted sources and baseline "facts" has shifted into an alternate reality. They no longer trust mainstream information channels ("enemy" sources are pre-emptively disqualified) and may have embraced a parallel narrative of events. In such a closed epistemic system, even impeccable logical evidence from outside is dismissed or reinterpreted as false, because it doesn't come from inside the bubble. *Tells:* The individual frequently claims that any counter-evidence is fake or comes from conspirators; they might assert that *nothing* from the opposing side can be believed (source gating), and they often propose radically different versions of basic facts (e.g., claiming widely documented events never happened or were staged).

Evidence: The concept of epistemic closure closely mirrors what researchers observe in **echo chambers and conspiracy subcultures**. When someone is deep in a conspiracy theory, their worldview is often self-sealing: **contradictory evidence is explained away as part of the conspiracy itself, which makes the belief system resistant to falsification** [29] . For example, if you present a QAnon believer with official records disproving one of their claims, they might say "Those records were fabricated by the deep state" – thereby nullifying the evidence by accusing the source. As a 2015 analysis by Lewandowsky et al. noted, conspiracy theories inherently *"postulate that conspirators use disinformation to cover up their actions – implying that people who try to debunk the conspiracy may themselves be part of it"* [29] . This is exactly an epistemic closure dynamic. The person has *redefined whom to trust* so any external correction is automatically suspect.

Likewise, research on *conspiratorial reasoning* finds that alternate belief systems can have **internally consistent but false premises that lead to extreme misperceptions**. Lewandowsky et al. (2013) pointed out that conspiracy believers can protect a *"cherished belief… by casting overwhelmingly disconfirmatory evidence… as the product of a conspiracy"* [30] . One outcome is that traditional debunking (which relies on common factual ground) fails because you don't even agree on what sources or institutions are legitimate. In misinformation studies, this is sometimes referred to as **source credibility gating** – e.g., hardcore partisans might reject a true news story purely because it appeared in the New York Times (a "hostile" outlet to them). A recent study in *Nature* found that labeling sources and providing social cues of credibility can help those *not* in closed bubbles discern truth, but has little effect on those who have **completely lost trust in mainstream media** [31] . They have effectively alternate trust anchors (maybe fringe blogs, partisan pundits, or "what my community says") and an alternate set of facts.

Distinctness: Epistemic closure is broader than the other factors and often the culmination of them. It differs from the simple belonging blocker in that it's not just *who* says it, but an entire *ecosystem* of accepted truths vs "forbidden" information. It's also not exactly the same as identity defense – it's more about a wholesale replacement of epistemology (how one determines truth). The thesis distinguishes it by focusing on **"source ecosystems and allowed premises."** Indeed, once someone's allowed premises diverge (e.g., they believe all scientific authorities are corrupt so any evidence from them is invalid), traditional logic doesn't work because you can't even agree on premises. Empirically, researchers Nguyen (2020) and others describe echo chambers where trust is only given to in-group voices, and all outsiders are distrusted *by default*.

A stark illustration is **the persistence of the false belief that the 2020 U.S. election was "stolen"**: believers consume only media that reinforce that narrative, and they dismiss contrary audit results or court rulings as lies by the corrupt establishment. Even showing them detailed evidence often hits a wall of "those

sources are all in on it." Thus, this blocker shows up as *total resistance unless the evidence can be presented from inside their trusted circle*. Experiments confirm this to a degree: if you provide the same factual correction via a channel the person trusts (say, a known conservative commentator correcting a conspiracy on a conservative platform), acceptance is much higher than if it comes from an official government or liberal media source [32] .

In summary, **epistemic closure is a fundamental barrier** because it means the person's **"gatekeeper" for reality has changed**. All external logic is filtered out, so one must first penetrate the bubble of trust. It stands as a distinct internal mechanism – essentially a reconfiguration of the person's knowledge network – and it's arguably the hardest to crack. Neutralizing it might require *internal sources* (members of the community) to carry the correction, since outside voices are muted. It's worth noting that epistemic closure often arises from sustained operation of other blockers (repetition, identity, moral polarization leading someone gradually into an alternate reality), but once in place, it has a self-sustaining quality. No ninth factor may be needed to explain further resistance – you have a closed loop.

Other Potential Blockers or Nuances Missing from the Thesis

The eight blockers above are each strongly supported by research and together cover a wide array of cognitive, social, and emotional reasons why people resist factual updates. The thesis boldly claims these eight are **exhaustive** and any observed resistance can be explained by one or more in this set. A critical evaluation should ask: *Is anything important missing?* Are there other independent psychological mechanisms, meeting the criteria, that could qualify as a ninth blocker? Here we consider a few possibilities raised in the broader literature:

- **Overconfidence and the Dunning–Kruger Effect:** Some individuals resist correction simply because they *overestimate their own knowledge or reasoning abilities*. This is the person who says, "I've done my research, I know better than the so-called experts." Overconfidence is an internal state that can make someone immune to evidence – they assume they can already discern truth perfectly, so any contrary info must be wrong. Research has found that **people who are most overconfident in their "fake news" detection skills are often *worse* at it and more likely to engage with misinformation** (a classic Dunning–Kruger pattern) [33] . Such individuals also **dismiss differing opinions and even expert advice, reinforcing their own echo chamber** [33] . This factor might not be fully captured by "authority cues" or "identity." It's more about meta-cognition – a poor self-awareness of one's ignorance. An intervention that specifically targets this (for example, showing people that they answered many factual questions wrong to puncture overconfidence) can make them more receptive to learning. So, overconfidence bias could be considered a distinct blocker, as it fulfills the criteria: internal (yes, it's in the mind), distinct (one can be overconfident regardless of group or moral factors), causally impactful (inflated self-assurance does correlate with resisting new info), detectable (e.g., by quizzing knowledge vs. confidence), and specifically targetable (through feedback or "calibration" training). The thesis did not explicitly list this, and it might be a noteworthy addition. However, one might argue it relates to Bandwidth (a person with low ability but high confidence is effectively not processing info properly) or Authority (they reject outside authority, trusting their own authority). Still, the Dunning–Kruger phenomenon is a specific cognitive bias worth acknowledging in the context of misinformation.

- **Psychological Reactance (Autonomy Defense):** Reactance is a well-established phenomenon where people resist persuasion if they feel their freedom to think or choose is being threatened. If a

corrective message comes off as too forceful, patronizing, or mandatory ("You *must* believe this!"), it can trigger an oppositional response: the person doubles down on the false belief just to reassert their freedom. This is an *affective* reaction rooted in personality and situational context. Studies in health communication show that **highly controlling or high-threat messages often backfire by motivating people to generate counter-arguments and resist the message** [34] . In the misinformation realm, the notorious "backfire effect" (while generally rare) can partly be attributed to reactance in certain cases – for instance, a correction delivered in a scolding tone by an authority might cause a prideful rejection rather than compliance. Reactance meets the criteria of an internal blocker: it's an internal motivational state ("don't tell me what to think!") distinct from the content of the belief itself. It can be isolated (e.g., keep content same but frame one message as a gentle suggestion and another as an authoritarian dictate – the latter is more likely to be rejected [34]). It's detectable through tells like expressions of anger at the source or explicit statements of refusing to be influenced. And it's intervention-specific: one can reduce reactance by using a different approach (e.g., *affirm the person's freedom* before presenting the correction, or use a question Socratically rather than a directive). The thesis did not list reactance as one of the eight, which might be a gap. Possibly, the author assumed reactance is part of Identity or Reputation (feeling one's self threatened), but reactance is a broader response that can occur even if no identity is implicated – it's about *control*.

- **"Existential" Motives – Fear and Control:** As noted in the literature on conspiracy theories, beyond the epistemic (truth) and social motives, people often have **existential motives** for holding certain false beliefs [35] . These include the desire to feel safe, in control, or to make sense of a scary world. Sometimes, accepting the *truth* is emotionally harder than clinging to the false belief. For example, climate change denial can be driven by the uncomfortable fear and helplessness the truth would induce – denial is a coping mechanism. Similarly, during the COVID-19 pandemic, some people believed misinformation about miracle cures or hoaxes because the reality of a deadly virus out of one's control was too frightening. These emotional coping needs can be an internal blocker: the correct information triggers anxiety or despair, so the person's psyche rejects it to protect itself (this is related to classic denial in psychology). The thesis's "Identity" category touches on worldview threat, but existential comfort might merit its own emphasis. It's distinct in that it's not about social belonging or morality, but about managing one's own fear or uncertainty. Indeed, a review on conspiracy beliefs concluded that *conspiracy theories may appeal because they promise to satisfy epistemic AND existential needs – like reducing uncertainty or providing a sense of control – even if ultimately they don't deliver true satisfaction* [35] [36] . A "fear-based" blocker would have tells such as apocalyptic or security-related themes in the person's discourse and might be isolated by varying whether the truth is scary or not. In practice, interventions like stress management or offering alternative sources of meaning might target this. While the eight-blocker framework doesn't explicitly list existential anxiety, it could be argued to fall under Identity (worldview) or Bandwidth (intolerance of ambiguity). But many experts consider it a key motive in its own right.

- **Confirmation Bias & Biased Reasoning:** One could point out that humans have a general tendency to confirm what they already believe, even without strong identity attachment. This is more of a generic cognitive bias: once we have an opinion, we selectively seek and interpret information to support it. The thesis's blockers like identity and familiarity cover aspects of this – indeed those give rise to confirmation bias. But plain confirmation bias can occur in trivial beliefs too (for instance, a person believes a brand of diet works and discounts any data showing it doesn't, simply because they've invested belief in it). This might be seen as a combination of cognitive inertia and dissonance

avoidance. It might not qualify as a separate "fundamental" mechanism here because it can be seen as an outcome of other drives (self-consistency, etc.). Still, it's worth noting that even absent strong emotions, the brain has a default to **prefer congenial information** and skepticism towards uncongenial information (as demonstrated in classic studies where people weighted evidence supporting their initial view as stronger than evidence against it). The thesis implicitly includes this in Identity Preservation and Epistemic Closure, so it's not truly missing, but it is a baseline tendency that underlies many blockers.

Are the Eight Truly Distinct and Sufficient? Based on our review, the eight factors cover the vast majority of internal influences documented by research. Cognitive limitations (item 1 and 7), social alignment (2 and 3), personal identity (4), heuristic cues (5 and 7), emotion/morale (6), and worldview closure (8) form a comprehensive taxonomy that maps onto what scholars often categorize as cognitive, social, and affective drivers [37] . There is some overlap inherently – e.g., moralized contempt often goes hand in hand with identity and belonging (group-based moral values), and epistemic closure often results from prolonged identity and repetition dynamics – but the thesis is aware of this and emphasizes isolating conditions. Empirical work also sometimes finds one factor dominating in contexts where others are held constant. For instance, in environments with ample cognitive resources and no social pressures, identity still may prevent change [13] ; or in low-identity, low-morale contexts, repetition alone can sway beliefs [25] . This suggests the list is hitting the main independent variables.

We did **not find clear evidence of a ninth fundamental blocker** that unquestionably meets the criteria without collapsing into one of the eight. The closest candidates we discussed (overconfidence, reactance, fear) are indeed important, but one might argue they are secondary contributors or nuances. Overconfidence could be seen as a facet of authority (self as authority) or just lack of metacognitive bandwidth. Reactance is a particular emotional reaction that often piggybacks on identity or reputation threats ("you're trying to control me" can be interpreted as a threat to autonomy, part of one's identity of being free, or simply personal resentment – it's a grey area). Existential fear motive often ties into identity (worldview about what is safe/controlled) or moral narratives ("the evil others are causing chaos, we must believe this to feel in control").

Crucially, we should ask: **Is there any documented instance of durable belief resistance that cannot be explained by some combination of these eight?** So far, the literature doesn't provide a clear counterexample. Many attempts to find a mysterious "backfire effect" or irrational stubbornness without a cause have ended up attributing it to one of these known biases (or finding that, once measured properly, people actually *do* update beliefs more often than assumed [38]). The thesis also sets a falsification criterion: if an intervention targeting one blocker in isolation had no effect when theory says it should, that would undermine the model. We haven't encountered a published result that cleanly violates those orthogonality expectations – though not all have been tested rigorously in combination.

One area of partial debate: Some studies (e.g., a 2021 study by Ecker et al.) found *no effect of partisan worldview on certain fact corrections* [39] , suggesting that sometimes identity doesn't always block updates if messages are crafted well. This doesn't disprove the existence of the blocker, but it reminds us these factors are context-dependent. Similarly, the so-called "backfire effect" (where corrections increase false beliefs) has proven extremely elusive in large samples [40] – it's not that people typically go in the opposite direction, they just often *don't change*. This implies that when corrections are failing, it's usually because one or more blockers are at work, not some inexplicable perversity.

In conclusion, the eight proposed fundamental blockers are strongly grounded in research. They capture the key internal dynamics that scholars have identified as obstacles to belief change in the face of evidence. A couple of additional nuances (overconfidence and reactance) deserve attention, but they may be folded into the broader categories or considered sub-blockers. There is **no obvious ninth mechanism** widely recognized in the literature that stands completely outside these eight. If anything, the thesis's comprehensive list is a strength – it integrates insights from cognitive psychology (e.g., fluency, dissonance), social psychology (identity, conformity), and communication (source credibility, misinformation studies). The challenge for practitioners is how to **neutralize multiple blockers simultaneously**, since in real-world scenarios several often operate together (e.g., a false belief might be appealing because it's repeated *and* identity-affirming *and* championed by one's group). But the thesis provides a framework for diagnosing which levers are in play.

Bottom line: The eight blockers thesis stands up well to a critical review – each blocker is supported by empirical evidence as a distinct factor in belief perseverance. While we can suggest minor additions or overlaps (and emphasize that human psychology is complex), the core claim that these are the fundamental internal reasons "why facts don't change our minds" is in line with current scientific understanding [1] [2]. No glaring "missing piece" emerged from our research, though continued studies might further refine the interactions between these factors. If all active blockers can truly be neutralized – increasing cognitive engagement, affirming identity, providing in-group messengers, easing fear, etc. – then indeed we would expect rational evidence to, at the very least, face **far less active resistance** from the mind. Whether that guarantees acceptance is debatable, but the theory posits it would, and so far, no evidence falsifies that optimism.

Sources

- Ecker, U. K. H., et al. (2022). *The psychological drivers of misinformation belief and its resistance to correction.* **Nature Reviews Psychology, 1**(1), 13–29. This review covers cognitive, social, and affective factors behind misinformation beliefs and difficulties in correcting them [1] [29].
- Wyer, N. (2010). *Selective self-categorization and the in-group persuasion effect.* **Journal of Social Psychology, 150**(5), 452–470. Demonstrates that people are more persuaded by in-group sources than out-group sources, confirming the role of group allegiance in acceptance of arguments [6].
- Cialdini, R. (various works summarized in blog). *On public commitments and resistance to persuasion.* Summarizes research that public commitments increase resistance to attitude change due to consistency and reputation concerns [8].
- Kahan, D. (2017). *Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition.* Cultural Cognition Working Paper. Synthesizes evidence that people dismiss evidence threatening to their cultural identities and that corrections are resisted when they challenge identity-affirming beliefs [12] [13].
- Brinol, P. & Petty, R. (2009). *Source factors in persuasion: A self-validation approach.* (Review). Discusses how credible sources can shortcut persuasion processes [16].
- Pretus, C., et al. (2023). *Neural and behavioral correlates of moral conviction.* (Referenced in Paris Institute article). Reports that moral conviction amplifies negative emotions (like hate) toward opponents and rigidity of attitudes [20].
- Pennycook, G., et al. (2018). *Prior exposure increases perceived accuracy of fake news.* **Journal of Experimental Psychology: General, 147**(12), 1865–1880. Empirical demonstration of the illusory truth effect in fake news context [25].

- Lewandowsky, S., et al. (2015). *"Conspiracist cognition" and information resistance.* (Inferred from references). Notes that conspiracy theories create a closed loop where disconfirming evidence is reinterpreted as part of the conspiracy [29] .
- Altay, S., et al. (2020). *Why do so few people share fake news? It hurts their reputation.* **New Media & Society, 24**(6), 1303–1324. Finds that people's reluctance to share misinformation is largely due to fear of reputational damage [10] .
- Additional references as cited inline above (e.g., studies on analytic thinking and fake news [3] [4] , need for closure and conspiracy belief [41] , etc.) which support specific points in the analysis.

---

[1] [3] [4] [15] [17] [18] [24] [25] [26] [27] [37] [38] [39] [40] The psychological drivers of misinformation belief and its resistance to correction | Nature Reviews Psychology

https://www.nature.com/articles/s44159-021-00006-y?
error=cookies_not_supported&code=2decf76a-3ee1-4fd2-8049-067130dd186b

[2] [29] [30] [32] [35] [36] [41] The Psychology of Conspiracy Theories - PMC

https://pmc.ncbi.nlm.nih.gov/articles/PMC5724570/

[5] AIS Electronic Library (AISeL) - AMCIS 2024 Proceedings: The Impact of Cognitive Load on Belief in Fake News and Real News

https://aisel.aisnet.org/amcis2024/security/security/4/

[6] [7] Selective self-categorization: meaningful categorization and the in-group persuasion effect - PubMed

https://pubmed.ncbi.nlm.nih.gov/21058574/

[8] [9] [11] Commitment & Consistency: Cialdini's Principle in Marketing

https://www.cognitigence.com/blog/commitment-and-consistency-principle

[10] (Why) Is Misinformation a Problem? - PMC - NIH

https://pmc.ncbi.nlm.nih.gov/articles/PMC10623619/

[12] [13] [14] Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition by Dan M. Kahan :: SSRN

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2973067

[16] Don't believe them! Reducing misinformation influence through …

https://pmc.ncbi.nlm.nih.gov/articles/PMC11345350/

[19] [20] [21] [22] [23] PPIAS - Proceedings of the Paris Institute for Advanced Study

https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence/

[28] Perceived source credibility mediates the effect of political bias on …

https://www.sciencedirect.com/science/article/abs/pii/S0191886921006486

[31] Source-credibility information and social norms improve truth …

https://www.nature.com/articles/s41598-024-57560-7

[33] How the Dunning-Kruger Effect Impacts Overconfidence - Helio

https://helio.app/ux-research/laws-of-ux/dunning-kruger-effect/

[34] Understanding Psychological Reactance: New Developments and …

https://pmc.ncbi.nlm.nih.gov/articles/PMC4675534/