

## REVIEW

## When identity meets rationality: The arousal of motivated System 2 reasoning

Hongyang Liu<sup>\*</sup>, Jana Kvintova<sup></sup>, Justyna Dockalova<sup></sup>, Lucie Vachova<sup></sup>,  
Irena Plevova<sup></sup>

Department of Psychology and Abnormal Psychology, Faculty of Education, Palacky University, Olomouc, Moravia, Czech Republic

## Abstract

**Background:** Classical dual-process theories assume that fast, intuitive System 1 often produces bias, whereas slow, deliberative System 2 corrects it. Evidence on motivated reasoning suggests a refinement: when beliefs are tied to identity, deliberation can intensify rather than reduce bias (motivated System 2 reasoning; MS2R).

**Objective:** To synthesize theory and evidence on when and why analytic reasoning amplifies identity-consistent judgments, and to highlight implications for health-related beliefs and misinformation.

**Methods:** Focused narrative review integrating findings from political, science-communication, and health literatures, including high-powered replications and resource-rational/approximate Bayesian models, to organize MS2R mechanisms, boundary conditions, and countermeasures. Identity-congruent evaluation of information is robust across domains (e.g., climate risk, vaccination, gun policy), whereas the strong claim that greater cognitive ability uniformly amplifies bias receives mixed support. We organized five proximate mechanisms that can yield MS2R: confirmation bias/motivated skepticism, cognitive dissonance rationalization, identity-protective cognition, expressive responding/instrumental rationality, and bounded resource-rational selective trust. MS2R is most likely when identity stakes are high, evidence is ambiguous, accuracy incentives are weak, affective threat gates processing, audiences are salient, and information ecologies encourage selective trust. In health psychology, these dynamics help explain the persistence of vaccine hesitancy and health misinformation despite deliberation. We outlined countermeasures that realign deliberation toward accuracy, including accuracy prompts/incentives, identity-safe and value-concordant framing (including self-affirmation), structured symmetric scrutiny, probabilistic forecasting with feedback, prebunking, and interventions that recalibrate source reliability.

**Conclusion:** System 2 is not inherently corrective: when identity stakes dominate, deliberation can intensify bias. Health interventions should reduce threat, align values, and incentivize accuracy to curb misinformation.

**Keywords:** Motivated reasoning; System 2, Dual-process theory; Identity-protective cognition; Expressive responding; Bounded/resource-rationality; Polarization; De-biasing

**\*Corresponding author:**  
Hongyang Liu  
(hongyang.liu@upol.cz)

**Citation:** Liu H, Kvintova J, Dockalova J, Vachova L, Plevova I. When identity meets rationality: The arousal of motivated System 2 reasoning. *Health Psychol Res.* 2026;14(2):0326.  
doi: 10.36922/hpr.0326

**Received:** October 9, 2025

**Revised:** January 5, 2026

**Accepted:** January 12, 2026

**Published online:** May 4, 2026

**Copyright:** © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons AttributionNon-Commercial 4.0 International (CC BY-NC 4.0), which permits all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Publisher's Note:** AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 1. Introduction

Dual-process accounts propose that human judgment relies on two modes of cognition. System 1 denotes fast, automatic, associative processing that relies on heuristics; System 2 denotes slower, effortful, and reflective processing that supports rule-based computation and abstract reasoning.<sup>1,2</sup> In the traditional view, many biases originate in System 1 and are corrected when System 2 engages to scrutinize initial impressions, yielding accurate beliefs and decisions.<sup>1,2</sup> This perspective undergirds common de-biasing strategies in education, science communication, and health counseling that emphasize critical thinking and numeracy.

Research on motivated reasoning, however, complicates this optimistic picture. Motivated reasoning refers to goal-directed evaluation and inference—often driven by identity, values, or preferences—wherein cognitive resources are selectively deployed to reach desired conclusions rather than accuracy-maximizing ones.<sup>3</sup> Classic studies show that politically sophisticated individuals can exhibit stronger confirmation bias and disconfirmation bias when evaluating mixed evidence on identity-relevant issues, becoming more—not less—polarized after deliberation.<sup>4</sup> In risk perception, greater science literacy and numeracy have been linked to larger ideological divides on contentious topics (e.g., climate change), an effect attributed to identity-protective cognition (IPC)—using one's best reasoning to defend group-congruent beliefs.<sup>5</sup>

These findings motivate a refinement to dual-process theory: Motivated System 2 reasoning (MS2R). MS2R posits that System 2 is functionally conditional. Under accuracy motivation, deliberation can correct intuitive errors; under identity or directional motivation, the same deliberative machinery can serve as a “defense attorney,” generating selective scrutiny of incongruent evidence and fluent rationalizations for congenial claims.<sup>3,6</sup> In short, analytic thinking is not uniformly de-biasing; when identity is implicated, it can amplify bias by enabling more sophisticated justification of preferred conclusions.<sup>4,7</sup>

This refinement is particularly important for health psychology. Numerous health-relevant beliefs and behaviors—vaccination decisions, perceptions of medical risk, adherence to preventive guidance, and nutrition or exercise norms—are entangled with social identities, moral values, and group norms. The misinformation literature shows that false health claims can persist despite corrective information, and corrections may be discounted when they threaten identity-relevant worldviews.<sup>8</sup> Moreover, recent experiments have shown that prompting deliberation or accuracy goals often improves truth discernment in

general, yet identity threat can still gate how evidence is processed, inviting selective counter-arguing rather than open-minded updating on charged topics.<sup>9,10</sup> Understanding when System 2 serves accuracy versus identity protection clarifies why information-heavy health interventions sometimes underperform and why framing, identity-safety, and incentive design can be decisive.

The present review contributes a novel integration and roadmap tailored to health psychology. First, key constructs are defined consistently: System 1 (automatic), System 2 (analytic), motivated reasoning (goal-consistent information processing), and IPC (deliberation in service of group-congruent beliefs).<sup>1–3,5</sup> Second, MS2R is explicitly positioned as a revision within dual-process theory, specifying boundary conditions under which System 2 shifts from de-biasing to rationalizing. Third, evidence across politics, science communication, and health is synthesized to show that the direction of System 2's effect depends on motivational states and task affordances (e.g., ambiguity, incentives, audience visibility).<sup>4,7,9,10</sup> Fourth, the review assembles complementary mechanisms—motivated skepticism, IPC, cognitive dissonance regulation, argumentative/persuasive functions of reasoning, and resource-rational selective trust—to explain MS2R's emergence and variability.<sup>11–13</sup> Finally, practical implications for health psychology are emphasized: identity-safe framing, self-affirmation, accuracy-goal prompts, symmetric scrutiny (e.g., steel-manning), probabilistic forecasting, and prebunking can realign deliberation toward accuracy in health contexts.<sup>8,9,11</sup>

The review proceeds from the traditional dual-process account to the MS2R revision; surveys empirical evidence across political, scientific, and health domains; organizes mechanisms that can produce MS2R; considers boundary conditions and critiques; and concludes with intervention strategies and research priorities directly applicable to health-related belief formation and behavior change.

## 2. Theoretical framework

### 2.1. Classical dual-process account and motivated System 2 reasoning

Psychologists have long believed that System 2 is the key to “de-biasing.” Because the intuitive judgments of System 1 rely on experience and heuristic strategies, System 1 often produces cognitive biases, such as confirmation bias, base-rate neglect, and availability bias. One function of System 2 is to monitor and correct these biases. When System 2 engages to scrutinize initial impressions, apply rules, and integrate evidence, it reduces bias and improves accuracy from System 1.<sup>14</sup>

For example, in the classic cognitive reflection test (CRT), intuition often yields a misleadingly “obvious” answer, whereas people with high cognitive reflection ability are able to activate System 2, suppress the intuitive impulse, and obtain the correct answer.<sup>15</sup> Studies show that individuals with high cognitive reflection ability perform better at avoiding common cognitive traps (such as base-rate neglect and logical errors), supporting the view that System 2 has a corrective function.<sup>15,16</sup> The implicit assumption of this traditional view is that most biases originate from the rapid processing of System 1, and as long as an individual’s System 2 is engaged in more thorough thinking, bias should be reduced.

As Sunstein<sup>17</sup> pointed out, if it is recognized that people’s thinking is influenced by heuristic-driven intuitions, improving individuals’ analytical thinking abilities (e.g., digital literacy and critical thinking) should help overcome those intuitive biases. In popular science and education, this idea is reflected in the so-called “deficit model”: it assumes public misunderstandings stem from insufficient knowledge or rationality, so providing more facts and cultivating more rational thinking should bridge the gap.<sup>18</sup>

Evidence on motivated reasoning, however, indicates that System 2’s corrective role is conditional. When accuracy goals dominate, and identity stakes are low, deliberation typically improves performance; when identity maintenance goals are salient—because beliefs are tied to group identity or moral commitments—System 2 may be recruited to defend prior commitments, a pattern described as MS2R.<sup>3,4,7</sup> Recent high-powered and preregistered tests have qualified earlier claims that cognitive sophistication necessarily amplifies partisan bias: several studies reported that higher numeracy improved accuracy without strengthening directional bias, suggesting that identity stakes, task features, and incentives govern whether System 2 acts as an auditor or an advocate.<sup>19–22</sup> Related evidence in science and climate domains similarly indicates that numeracy often tracks knowledge gains that can reduce bias once topic knowledge is taken into account, reinforcing the view that boundary conditions—rather than ability alone—determine System 2’s function.<sup>23</sup>

## 2.2. Instrumental rationality and incentives

Instrumental rationality may involve the following pathways: One is IPC, an internal motive to protect self-concept and group identity by interpreting information in identity-congruent ways, often via selective attention, counter-arguing of dissonant claims, and rationalization to reduce threat.<sup>3,5,7</sup> A second pathway is expressive responding (value-signaling), an external and instrumental motive to signal loyalty or values to relevant audiences

for reputational or coalition benefits, even when private beliefs are ambivalent.<sup>24,25</sup> Both pathways can yield identity-congruent responses, but they differ in proximate motivated defensive self-protection versus strategic signaling—implying different intervention levers.

These pathways dovetail with classic cognitive mechanisms. Confirmation bias privileges congenial information in seeking, interpretation, and memory—typically an upstream, evidence-handling process—whereas cognitive dissonance refers to the aversive arousal from inconsistency that motivates post-hoc rationalization or attitude change downstream.<sup>4,26–28</sup>

Contemporary resource-rational perspectives further highlight that bias can enter upstream via credibility judgments about whom to believe before what to believe. Under cognitive and social constraints, prioritizing in-group or otherwise “trusted” sources may be instrumentally efficient, yet it canalizes System 2 toward identity-congruent evidence streams.<sup>29–31</sup> Neurocognitive findings are consistent with this pipeline: politically charged evaluations engage networks linked to valuation, error monitoring, and mentalizing, and identity-congruent versus incongruent sources elicit differential responses in these systems, aligning with selective-trust plus motivated evaluation.<sup>32</sup>

These clarifications are consequential in health psychology, where beliefs about vaccination, medical risk, prevention, and lifestyle often implicate social identities and community norms. When identity stakes exceed accuracy goals, System 2 can shift from correcting myths to defending identity-congruent narratives unless interventions also reduce identity threat, alter reputational incentives, and leverage credible sources. Evidence indicates that brief accuracy prompts reliably improve truth discernment and reduce the willingness to share misinformation, including in health contexts.<sup>33,34</sup> Complementary psychological inoculation (prebunking)—short games or videos teaching manipulation tactics—can build resilience to misinformation, with newer work delineating durability limits and memory-based “booster shots” that extend protection over time.<sup>35–38</sup> Together, these findings support a threshold model: System 2 tends to de-bias when accuracy goals are salient, and identity stakes are managed, but can rationalize when identity goals dominate; effective interventions therefore target both motivation (identity safety, incentives, reputational context) and cognition (attention to accuracy, prebunking, and structured evaluation).

Finally, although our focus is on proximate psychological mechanisms and incentives, a brief ultimate-level backdrop helps contextualize why deliberation can so

readily take an argumentative form in identity-relevant settings. As an ultimate-level backdrop, the Argumentative Theory of Reasoning proposes that human reasoning evolved primarily to produce and evaluate arguments in social contexts.<sup>11</sup> From this perspective, deliberation is not inherently truth-tracking and can be naturally oriented toward justification and persuasion, which helps contextualize why analytic thinking may sometimes amplify identity-consistent conclusions.

### 3. Empirical evidence review: When does System 2 correct VS. defend

#### 3.1. Directional motivated reasoning across political and scientific domains

Empirical research across politics, science, and health generally supports the claim that identity involvement—rather than topic domain per se—plays a central role in determining when System 2 is used to correct versus defend biases. Classic work on motivated reasoning showed that people preferentially accept congenial evidence and scrutinize uncongenial evidence, thereby polarizing attitudes even when both sides receive the same information.<sup>3,4,39</sup> More recent studies extend this pattern by examining how cognitive sophistication, numeracy, and science literacy interact with identity-relevant topics. Overall, a consistent directional motivated reasoning effect is observed—identity-congruent claims are easier to accept than identity-incongruent claims—but evidence is mixed regarding whether higher analytical ability reliably amplifies this bias.<sup>5,19,20,22</sup>

In the science and climate domain, Kahan *et al.*<sup>5</sup> demonstrated that as science literacy and numeracy increased, the perceived risk of climate change diverged rather than converged between liberals and conservatives. Individuals high in science literacy were more likely to align their risk judgments with their political identity, contrary to the knowledge-deficit expectation that more knowledge should uniformly reduce misperceptions. Related work on science beliefs shows similar identity-congruent polarization for issues such as nuclear power, fracking, or gun control.<sup>40,41</sup> However, newer evidence indicates that this pattern is not unconditional. For example, Cho *et al.*<sup>23</sup> reported that numerate individuals were more knowledgeable about climate science and, once knowledge was controlled, numeracy no longer increased polarization. Likewise, large-scale replications find strong main effects of identity congruence but fail to detect consistent numeracy  $\times$  identity interactions, suggesting that higher numeracy often improves overall accuracy without uniquely increasing partisan bias.<sup>19,20,22</sup>

The “motivated numeracy” paradigm illustrates both the strength and the limits of the MS2R claim. In Kahan *et al.*’s<sup>7</sup> study, participants interpreted a contingency table describing whether a new skin-care ointment improved rash. When the table concerned a neutral topic (skin cream), higher numeracy improved accuracy regardless of the correct answer. However, when the same numerical structure was framed as a politically charged question about whether gun control reduces crime, high-numeracy participants performed best only when the numerically correct answer aligned with their political identity; when the correct answer contradicted their identity, the advantage of numeracy disappeared or reversed. This pattern suggested that identity-protective goals can steer System 2 toward selectively using numeracy to rationalize congenial conclusions. Subsequent preregistered replications and extensions, however, have produced more nuanced results. Several studies replicate the basic directional congruence effect but not a robust amplification by numeracy<sup>19,20</sup>, while others show that cognitive sophistication primarily improves performance even on politicized items.<sup>22</sup> Together, these findings indicate that MS2R is not a universal outcome of high ability; it appears conditional on how strongly the task implicates group identities, how ambiguous the evidence is, and whether accuracy incentives are salient.

#### 3.2. Political and policy domains: Sophistication, elite cues, and polarization

Evidence from other political and policy domains converges on the same picture. Research on climate policy, partisan news, and judicial evaluations finds that people with greater political knowledge or ideological sophistication are often more consistent and more extreme in their partisan alignments, especially when elites provide clear cues.<sup>40,42–44</sup> In such contexts, more sophisticated citizens may be better at identifying plausible arguments supporting their side, leading to stronger attitude polarization even while their factual knowledge is higher. Yet this is not simply a matter of “smarter is more biased.” In several studies, expertise and numeracy enable accurate interpretation of complex information regardless of identity, particularly when tasks are transparently factual and identity stakes are low.<sup>23,29</sup> Taken together, the cross-domain evidence suggests a robust identity-congruence effect in evaluations but a contingent relationship between cognitive sophistication and bias: analytical ability can either mitigate or magnify bias depending on how tightly beliefs are bound to identity and how the task is structured.

#### 3.3. Health-related beliefs, misinformation, and accuracy intervebtions

Health-related domains show a similar interplay

between identity and deliberation. During the COVID-19 pandemic, several studies documented that belief in false headlines or misinformation about the virus was lower among individuals who engaged in more analytic thinking or were prompted to attend to accuracy.<sup>9,10,33,45</sup> At the same time, acceptance of vaccine misinformation and conspiracy narratives was strongly associated with partisan identity and distrust of out-group sources, even among highly numerate or educated individuals.<sup>30,31,46</sup> Studies on numeracy and medical risk communication likewise find that higher numeracy often improves comprehension of statistical risk, yet identity-linked beliefs about vaccination, preventive behaviors, or alternative medicine can still guide how those statistics are interpreted and which sources are trusted.<sup>23,46</sup> Thus, in health contexts, analytical skills tend to increase understanding of evidence, but identity-protective motivations and selective trust can channel System 2 toward defending pre-existing attitudes.

A growing line of research explicitly compares trait-level differences with situational manipulations. Trait measures, such as numeracy, the CRT, or open-minded thinking, reliably predict better performance on numerous reasoning tasks and reduced susceptibility to generic misinformation.<sup>10,47</sup> Yet on politicized topics, the same traits sometimes correlate with consistent identity-congruent reasoning, especially when individuals are highly engaged and perceive the issue as central to their values.<sup>5,40,43</sup> Situational manipulations reveal that motivated reasoning is not simply a function of slow versus fast thinking: people can produce identity-congruent judgments even under time pressure, and accuracy prompts or incentives often reduce misperceptions by shifting goals rather than by changing sheer processing depth.<sup>9,21,33,45</sup> This pattern supports the view that MS2R depends jointly on who is reasoning (traits and prior knowledge), what is at stake (identity involvement), and how the reasoning context is framed (goals, incentives, and task design).

### 3.4. Neural evidence and interim summary

Neuroscientific studies further illustrate how identity involvement shapes deliberation, while also underscoring the need for interpretive caution. For instance, politically charged reasoning has been associated with activation in regions involved in conflict monitoring, valuation, and emotion (e.g., anterior cingulate cortex, ventromedial prefrontal cortex, striatum), and identity-congruent versus incongruent arguments sometimes elicit distinct neural responses.<sup>13,32</sup> These patterns have been interpreted as neural correlations of cognitive dissonance or motivated evaluation, consistent with the idea that System 2 can be harnessed to rationalize identity-protective responses. However, such studies are correlational and typically do

not distinguish decisively between emotional, strategic, and resource-rational explanations. They nonetheless align with behavioral evidence suggesting that, when identities are implicated, deliberation is often recruited to justify rather than to revise prior commitments.

In summary, across politics, science, and health, empirical findings converge on a coherent picture. Directional motivated reasoning—preferential treatment of identity-congruent information—is robust, and it appears across a wide range of topics and methodologies.<sup>4,5,31,40</sup> In contrast, the strong version of the MS2R claim—that higher cognitive sophistication generally increases bias—receives mixed support. High numeracy and analytical thinking often improve accuracy and reduce generic susceptibility to misinformation, but under conditions of strong identity involvement and ambiguous evidence, those same capacities can be used to selectively bolster pre-existing beliefs.<sup>7,19,20,22</sup> These patterns suggest that the critical moderator is the alignment of identity goals and accuracy goals: when accuracy and identity point in the same direction, System 2 helps; when they conflict, System 2 may be deployed as a skilled advocate. This raises a central question for the remainder of the paper: through which psychological mechanisms and incentives does identity involvement redirect analytical reasoning from correction toward justification? The next section examines these mechanisms in detail.

## 4. Mechanisms underlying increased bias by high cognizers on identity-related issues

Highly analytic individuals are not immune to bias; when identity or values are implicated, their System 2 reasoning can be recruited to defend rather than correct prior commitments. Five mechanisms have been proposed to explain why greater cognitive sophistication can sometimes exacerbate bias on identity-laden topics: (i) confirmation bias and motivated skepticism, (ii) cognitive dissonance and conflict rationalization, (iii) IPC, (iv) belief as value-signaling and expressive responding (including instrumental rationality), and (v) bounded/resource-rational approaches that yield selective trust. These mechanisms operate at different levels (cognitive, affective, social, and evolutionary), but converge on the idea that deliberation is often guided by directional motives. The subsections below clarify each mechanism, highlight what distinguishes it, and indicate how they jointly contribute to MS2R.

### 4.1. Confirmation bias and motivated skepticism

Confirmation bias refers to the tendency to preferentially seek, interpret, and remember information that supports

existing beliefs while neglecting or discounting information that contradicts them.<sup>28</sup> In identity-relevant settings, this often manifests as motivated skepticism: congenial evidence is accepted with minimal scrutiny, whereas uncongenial evidence is subjected to intense critical evaluation.<sup>3,4,24,40</sup> This is primarily a cognitive mechanism that operates at the level of information handling.

Classic demonstrations show that when individuals with strong prior attitudes evaluate mixed evidence, they judge pro-attitudinal studies as more methodologically sound and persuasive than counter-attitudinal studies, even when the objective quality is matched.<sup>4,39</sup> In such contexts, individuals with higher political knowledge or cognitive sophistication are often associated with more asymmetric processing, because they have more argumentative resources to critique threatening information and to bolster supportive information.<sup>24,40</sup> Rather than eliminating bias, analytic capacity is used to construct and evaluate evidence in a selectively favorable way, producing greater polarization among the “best informed.”

Recent formal models show that such patterns can even emerge from approximate Bayesian reasoning when people update both on content and on perceived source reliability.<sup>48</sup> Under uncertainty, it can be computationally efficient to interpret disconfirming evidence as a sign that the source is unreliable rather than that the prior belief is wrong, especially when the prior belief is anchored in previously trusted sources. Over time, this leads to selective assimilation of congenial information and systematic doubt toward uncongenial claims pattern empirically observed in both political and health-risk communication.<sup>31,46</sup> In MS2R terms, confirmation bias and motivated skepticism explain how System 2 can be deployed to filter evidence asymmetrically: the more analytic the thinker, the more sophisticated the selective scrutiny can become.

Importantly, confirmation bias is conceptually distinct from cognitive dissonance. Confirmation bias concerns how evidence is searched for, evaluated, and remembered. Cognitive dissonance, in contrast, concerns why beliefs or interpretations change (or rationalize) after conflict is experienced. Both processes can lead to resistance to challenging information, and they often overlap in biased evaluation; thus, we distinguish them primarily by underlying motives and possible outcomes: confirmation bias is typically outcome-directional and tends to preserve or reinforce prior beliefs, whereas dissonance reduction is conflict-directional and can be resolved either via rationalization or, in principle, by belief change.

#### **4.2. Cognitive dissonance and conflict rationalization: Affective pressure to restore coherence**

Cognitive dissonance theory holds that holding inconsistent cognitions, or a mismatch between cognition and behavior, produces an aversive state of psychological tension that individuals are motivated to reduce.<sup>26,27</sup> In identity-relevant domains, dissonance arises when new information conflicts with cherished beliefs or with publicly chosen positions. The resolution often involves rationalization: reinterpreting the evidence, re-weighting its importance, or adjusting attitudes to restore subjective coherence.

Neurocognitive studies indicate that dissonance and conflict resolution are associated with activation in regions, such as the dorsal anterior cingulate cortex (dACC) and anterior insula, as well as ventromedial prefrontal and striatal regions implicated in valuation and reward.<sup>12,13</sup> For example, when individuals advocate a position contrary to their prior attitudes, greater dACC and insula activation predict subsequent attitude change, suggesting that the brain detects the conflict and then adjusts beliefs to relieve discomfort.<sup>12</sup> In partisan reasoning, exposure to counter-attitudinal information can trigger conflict and negative affect, followed by increased activity in regions associated with emotion regulation and reward once a self-serving reinterpretation is reached.<sup>13</sup>

Cognitively sophisticated individuals may be especially capable of conflict rationalization: they can generate complex narratives and justifications that reconcile dissonant information with existing commitments. Rather than updating beliefs toward the data, System 2 is used to craft explanations that make the discrepancy appear less problematic (for instance, by questioning the motives of the source, redefining the issue, or shifting comparison standards). Conceptually, this mechanism partly overlaps with confirmation bias, which is widely understood to include biased interpretation and evaluation of information, not merely selective exposure. We therefore distinguish the two constructs primarily in terms of underlying motives and possible outcomes: confirmation bias is generally outcome-directional, tending to preserve or reinforce existing beliefs, whereas dissonance reduction is motivated by conflict reduction and can, at least in principle, be resolved either by rationalization or by belief change. Together, they can form a powerful combination: identity-threatening information is both less likely to be accepted and, when encountered, more likely to be reinterpreted in ways that restore subjective coherence.

### 4.3. Identity-protective cognition: Defensive use of System 2 to safeguard group identity

Identity-protective cognition refers to the tendency to use reasoning in ways that protect one's social identity, group standing, or self-concept, especially on issues that signal group membership or moral allegiance.<sup>5,6</sup> The core motivation is internal and defensive: maintaining a coherent self-image and avoiding social or psychological costs associated with appearing disloyal or mistaken in front of valued in-groups.

In IPC, individuals use analytic skills selectively to support identity-congruent conclusions and to counter identity-threatening evidence. This has been documented in domains such as climate change, gun control, and public health. For example, in the “motivated numeracy” paradigm, highly numerate participants correctly interpret a contingency table when it describes an identity-neutral scenario (e.g., skin-rash treatment), but systematically misinterpret the same numbers when the correct conclusion conflicts with their partisan position.<sup>7</sup> The pattern indicates that numeracy is not uniformly deployed; it is activated in ways that protect identity.

Neuroscientific findings align with IPC as a mechanism that recruits both effective and valuation systems. Identity-congruent information tends to elicit positive valuation signals, while identity-threatening information elicits conflict and a negative effect, followed by rationalization.<sup>13,32</sup> IPC, therefore, acts as an affect-first gate that shapes how System 2 is applied: analytic thinking is directed toward defending identity rather than neutrally assessing evidence. In health contexts, this can explain why individuals with high health literacy or numeracy may still reject vaccines or discount medical consensus when such positions are tightly bound to political, religious, or community identities.<sup>31,46</sup>

Identity-protective cognition is closely related to, but distinct from, expressive responding. IPC primarily concerns sincere belief formation and maintenance under identity threat, even in private. Expressive responding, in contrast, emphasizes the strategic expression of identity-congruent claims for reputational or social payoffs, as discussed next.

### 4.4. Belief as value-signaling, expressive responding, and instrumental rationality

Beyond defensive identity protection, beliefs and expressed judgments also serve expressive and instrumental functions. In political science and economics, the concept of expressive utility captures the idea that people derive psychological or social benefits from expressing identity-

congruent views, regardless of their truth value.<sup>49,50</sup> Expressive responding occurs when individuals give survey answers or public statements that signal loyalty or group values, even if those answers do not fully reflect private beliefs.<sup>24,25,51</sup> Here, the motive is external and strategic: maintaining reputation, solidarity, or status within a group.

From this perspective, using System 2 to produce identity-congruent justifications can be instrumentally rational. It maximizes payoffs in an “identity economy” where being seen as loyal to one's side brings reputational, emotional, or material benefits.<sup>49,50</sup> For political elites, professionals, or activists, such incentives are especially transparent: careers, donations, and audience support often depend on staying aligned with coalition narratives.<sup>42,52,53</sup> In these cases, analytic reasoning is not misfiring; it is precisely tracking the agent's goal structure. The cost of acknowledging uncomfortable facts may be high, while the cost of holding an inaccurate belief on a distant policy issue is low, so System 2 is directed toward constructing defensible, identity-congruent positions.

Empirically, expressive mechanisms are suggested when partisan gaps shrink markedly once accuracy is incentivized or responses are made anonymous. For instance, when participants are paid for correct answers or told their responses will be privately checked, partisan differences on factual questions often decline, indicating that at least part of the earlier divergence reflected expressive responding rather than deep-seated belief.<sup>9,10,25,51</sup> In MS2R terms, expressive responding and instrumental rationality broaden the account beyond internal bias: System 2 is recruited to achieve social and reputational goals, not only to manage internal consistency. This mechanism overlaps with IPC but remains conceptually distinct, because it emphasizes outward signaling and payoff structures.

Motivated System 2 reasoning partially overlaps with expressive utility theory but is not reducible to it. Expressive utility focuses on the satisfaction gained from expressing certain positions; MS2R more broadly concerns how analytic processing is engaged under directional motives, whether the motive is private (identity preservation, dissonance reduction) or public (signaling, reputation).

### 4.5. Bounded and resource-rational approaches: Selective trust as efficient but polarizing

Resource-rational and bounded rationality perspectives reinterpret some apparent “biases” as adaptive responses to limited cognitive resources and noisy environments.<sup>29</sup> In this view, the human mind approximates Bayesian inference but must rely on heuristics and simplifying assumptions to remain tractable. One example is the Bayesian Inference with an Independence Approximation

and Source Reliability (BIASR) model, which shows that agents who simultaneously update on content and source reliability, while assuming partial independence among pieces of evidence, can exhibit confirmation-like patterns even when trying to be rational.<sup>48</sup>

Under such conditions, discounting strongly disconfirming evidence from unfamiliar or out-group sources can be a resource-efficient heuristic. Rather than constantly reconsidering deeply entrenched beliefs, individuals may treat surprising information as a cue that the source is unreliable. Crucially, this form of source skepticism is typically provisional rather than rigid: trust assignments can be revised when counterevidence accumulates or when feedback makes differential source reliability salient. Consistent with this view, in dynamic settings with repeated exposure to diagnostic evidence about relative source accuracy, initial ingroup-biased trust can be substantially attenuated or even eliminated.<sup>54</sup> Early in learning, this can produce selective trust in in-group or previously reliable sources and skepticism toward out-group or discrepant sources.<sup>55,56</sup> However, when environments provide sustained diagnostic feedback about differential reliability, these trust weights can update, and initial ingroup biases may diminish or disappear.<sup>54</sup> In polarized information environments, where different groups inhabit distinct media ecosystems, this selective trust can yield sharply divergent belief systems that are internally coherent and, given each group's informational input, locally rational.

These models offer a distinct perspective on MS2R. While IPC and expressive responding emphasize motivational motives, resource-rational models emphasize computational constraints and source-credibility inferences. High-cognition individuals may be particularly adept at implementing selective-trust strategies: they can quickly identify seemingly plausible reasons to downgrade the credibility of dissonant sources (e.g., methodological criticisms, alleged bias), thereby preserving prior beliefs in a way that appears rational from within their information network. In health-related settings, this dynamic is evident when individuals accept complex statistical arguments from in-group alternative-medicine influencers while dismissing epidemiological evidence from mainstream public health institutions as untrustworthy.<sup>31,46</sup>

Resource-rational accounts, therefore, do not deny MS2R; instead, they show how motivated use of System 2 can be understood as optimizing for goals other than global accuracy (such as coherence with trusted beliefs, efficient use of limited attention, and protection from deceptive sources). This helps explain why simply urging people to “be more rational” is insufficient: from within their own

perceived evidence ecology, they already are.

#### **4.6. Integrating mechanisms and positioning motivated System 2 reasoning among related theories**

These mechanisms are not mutually exclusive and do not form a strict linear sequence; rather, they operate at different levels and timescales and often interact. A plausible process in many identity-relevant situations is the following: identity-threatening information triggers affective conflict and dissonance (cognitive dissonance, affect-first gating), which in turn activates IPC; System 2 is then recruited to implement confirmation bias and motivated skepticism in handling new evidence; expressive and instrumental considerations shape what is publicly stated and how strongly a position is defended; all of this takes place within an environment where bounded, resource-rational strategies favor selective trust in in-group sources over out-group sources.

In this sense, the mechanisms are complementary rather than competing. Confirmation bias and motivated skepticism describe how evidence is processed; cognitive dissonance and conflict rationalization describe the affective drive to resolve inconsistency; IPC and expressive responding specify identity-related motives that direct System 2; resource-rational and BIASR-type models specify computational constraints and source-credibility inferences. Several mechanisms can also be seen as more hierarchical or meta-level: resource-rational frameworks operate at a higher level of abstraction, under which more specific cognitive and affective processes unfold.

Clarifying how MS2R relates to neighboring theories is important for theoretical precision. Traditional dual-process theories assume that System 2 is generally de-biasing; MS2R revises this assumption by highlighting conditions under which System 2 amplifies bias due to identity and motivational influences. Expressive utility theory focuses on the satisfaction or payoff from expressing certain positions; MS2R incorporates expressive responding as one pathway but also encompasses private, non-expressive processes, such as IPC and dissonance-driven rationalization. MS2R fills this gap by identifying identity involvement, task ambiguity, incentives, and information ecology as key moderators. Bounded/resource-rational models, finally, show that MS2R-like patterns can be locally rational given individuals' goals and constraints, suggesting that effective interventions must change perceived goals, incentives, and trust structures rather than merely increasing analytical capacity.

For health psychology and related applied fields, this integrated framework implies that high cognitive ability is a double-edged sword. When accuracy motives are aligned with identity and incentives, System 2 supports

error detection and informed decision-making. When identity stakes, expressive payoffs, and selective trust push in the opposite direction, System 2 can become a highly capable advocate for misinformed or polarized positions. Subsequent sections on criticism, boundary conditions, and countermeasures build on this integrated mechanism framework.

## 5. Criticism of motivated System 2 reasoning

### 5.1. Replication, scope, and what “System 2” means

A balanced reading of the literature indicates that MS2R is not ubiquitous. Several high-powered replications and re-analyses suggest that early demonstrations may have overstated the extent to which greater cognitive sophistication reliably magnifies bias on identity-relevant topics. Large, preregistered studies revisiting “motivated numeracy” typically find robust directional motivated reasoning—identity-congruent answers are easier to endorse—but do not consistently observe an interaction with ability. In a nationally representative United States sample, Stagnaro *et al.*<sup>20</sup> reported that numerical ability improved overall accuracy across five contentious topics, whereas identity-congruence exerted an additive effect for all participants, with no interaction between numeracy and congruence. Similarly, Tappin *et al.*<sup>22</sup> observed that cognitive sophistication predicted better performance even on politicized items, and the pattern that “the smartest are most biased” often failed to replicate. In this section, we treat MS2R as arising from the interaction between stable cognitive or identity factors and context-dependent cues that trigger identity-relevant deployment of deliberation.

These findings highlight the importance of operationalizing System 2. Numerous demonstrations treat numeracy or CRT scores as direct proxies for System-2 engagement. Yet ability, disposition, and momentary engagement are distinct constructs. Numeracy reflects domain-specific computational skill; CRT scores capture a disposition toward reflective thinking; and state deliberation concerns the actual depth of analytic processing on a given trial.<sup>47,57,58</sup> If participants rarely engage in sustained deliberation when identity is threatened, numeracy scores may be poor stand-ins for genuine System-2 processing in those moments. Recent work on fast versus slow motivated reasoning further underscores that identity-consistent judgments can arise rapidly and are not always amplified when more time or opportunity to think is provided.<sup>21</sup>

Design features and motivational cues also shape whether MS2R emerges. Ambiguous data presentations, low stakes, and weak salience of accuracy can encourage shallow or identity-driven processing, whereas explicit accuracy prompts and incentives tend to reduce both

errors and partisan gaps.<sup>9,45</sup> When participants receive monetary bonuses for correct answers or are instructed to prioritize accuracy, polarization often shrinks, and overall performance improves, suggesting that some MS2R effects may reflect insufficient or misdirected deliberation rather than deliberation inherently gone awry.<sup>9,10</sup> Together, these findings motivate a more circumscribed claim: MS2R appears most likely when: (i) identity stakes are high, (ii) evidence is ambiguous or open to multiple interpretations, and (iii) accuracy goals are weak or diffuse.

Taken together, the evidence favors an interactionist interpretation of MS2R. Situational cues—such as identity threat or salience, evidential ambiguity, and accuracy incentives—primarily determine whether analytic reasoning is deployed for accuracy or for identity defense. Stable individual differences then shape the strength and sophistication of that deployment: analytic ability and reflective disposition provide the cognitive tools for generating and evaluating arguments, whereas identity investment and domain knowledge increase the motivation and capacity to defend identity-relevant beliefs. Thus, MS2R is best conceptualized as a state-dependent strategy whose likelihood and magnitude are moderated by trait-level resources and commitments, rather than as a purely situational effect or a fixed individual difference.<sup>4,9,10,47,57,58</sup>

Clarifying the theoretical scope of MS2R, therefore, requires specifying both (i) the situational triggers that recruit deliberation for identity defense and (ii) the trait-level resources and identity commitments that moderate how strongly this recruitment shapes judgments. When identity motivations are minimal or aligned with accuracy, for example, in identity-neutral risk judgments, technical domains where professional norms emphasize truth, or educational contexts that reward calibration, deliberation generally improves accuracy and reduces classic biases.<sup>10,23,47</sup> MS2R applies more narrowly to contexts in which identity-maintenance goals conflict with accuracy goals and where task affordances provide room for selective interpretation. Under such conditions, System 2 can act as a “defense attorney,” but it does so unevenly across topics, individuals, and designs.

### 5.2. Affect-first gating, neural evidence, and methodological fragility

A second line of critique emphasizes affective “gating” as a boundary condition on MS2R. Classic work on hot cognition and the effect heuristic shows that automatic evaluative reactions to political symbols and identity cues arise within milliseconds and bias subsequent processing.<sup>59,60</sup> Neuroimaging studies indicate that exposure to identity-incongruent or dissonant information engages regions

linked to conflict detection and aversive arousal, such as dACC and anterior insula, as well as valuation and self-relevance regions, including the ventromedial prefrontal cortex.<sup>12,13</sup> These signals appear before or in parallel with activity in control-related networks, suggesting that affective and self-relevant appraisals can shape the input to subsequent deliberation.

Threats to self-integrity or group status further heighten defensive processing. Experiments show that threatening a valued aspect of the self increases derogation of dissonant sources, whereas self-affirmation interventions can reduce defensiveness and promote more balanced evaluation of counter-attitudinal evidence.<sup>61–63</sup> Individuals with high identity investment—such as those who are politically knowledgeable or strongly identified with a party—often display more polarized, directional processing under such threat conditions.<sup>4</sup> From an MS2R perspective, these findings can be interpreted as evidence that identity-related affective responses “gate” the operation of System 2: analytic resources are engaged, but often only after the affective system has flagged certain interpretations as identity-threatening or identity-congruent.

At the same time, interpretations of neural evidence require caution. Correlations between activation in dACC, insula, or vmPFC and partisan reasoning do not uniquely differentiate among competing mechanisms, such as dissonance reduction, expressive responding, or resource-rational selective trust.<sup>12,13,32</sup> Neuroimaging studies rarely manipulate identity stakes, incentives, and ambiguity orthogonally, making it difficult to determine whether System 2 is best characterized as “hijacked” by identity or simply following context-sensitive utility functions. Consequently, neural findings are better viewed as convergent but non-definitive support for the notion that identity involvement shapes deliberation, rather than as direct proof that reasoning itself becomes inherently more biased with ability.

Methodological critiques further highlight how fragile some MS2R effects can be. Small changes in framing (e.g., labeling identical numerical problems as neutral vs. politicized), item difficulty, or identity cues can shift whether numeracy helps or harms accuracy, and some hallmark interactions disappear under preregistered protocols or stronger accuracy incentives.<sup>7,19,22</sup> These patterns suggest that MS2R should be conceptualized as contingent on affective gating and task design, rather than as a general property of System-2 engagement.

### 5.3. Expressive responding, instrumental rationality, and boundary conditions

A third critique reframes various identity-congruent

responses as instrumentally rational signals rather than purely cognitive errors. Research on expressive responding indicates that partisans sometimes treat factual survey questions as opportunities to “cheerlead” for their side or to derogate opponents. When responses are anonymous or when monetary incentives are offered for correct answers, partisan gaps in reported beliefs often shrink substantially, implying that a portion of observed divergence reflects strategic expression rather than deeply held misperceptions.<sup>9,25,51</sup> In this view, at least some instances of MS2R arise because System 2 is used to maximize expressive utility—to signal loyalty, maintain reputation, or avoid social sanctions—not because individuals are genuinely convinced by their own arguments.

Expressive and instrumental motives are especially salient for elites and political professionals. Models in political economy describe office-seekers as utility maximizers whose careers depend on maintaining support among donors, primary voters, and ideological activists.<sup>52,53,64</sup> Since these constituencies are often more extreme than the mass public, deviating from coalition orthodoxy on identity-charged issues can entail real costs. Under such conditions, producing sophisticated, identity-consistent arguments can be career-rational even when decision-makers privately recognize countervailing evidence. Social pressures, such as preference falsification and the “spiral of silence,” further discourage public expression of dissenting views, reinforcing incentives to align expressed reasoning with group narratives.<sup>65,66</sup>

Alternative explanations also emphasize the need to distinguish between identity-colored expression and genuine belief distortion. Some critics argue that respondents may not literally believe extreme partisan claims but endorse them as a form of loyal signaling—captured in colloquial phrases such as “they might be a liar, but they are our liar.”<sup>31</sup> Experimental designs that combine accuracy incentives with private response formats suggest a mixed picture. Partisan gaps typically narrow but do not vanish under incentivized truth conditions, indicating that expressive responding explains part, but not all, of the bias.<sup>9,10,25,51</sup> This pattern is consistent with an interpretation in which MS2R encompasses both sincere IPC and strategic expressive responding, with their relative contributions varying across domains, audiences, and stakes.

These critiques, taken together, call for a refined conception of MS2R. Rather than positing that greater cognitive ability inherently produces greater bias, a boundary-conditions account is more defensible. System 2 tends to reduce error when: (i) identity stakes are low or aligned with accuracy, (ii) evidence is relatively clear, and (iii) incentives and norms reward correctness. In contrast,

MS2R is most likely when identity or reputational goals conflict with accuracy goals, ambiguity provides latitude for motivated interpretation, affective reactions gate processing, and expressive or instrumental payoffs favor identity-congruent signaling. Under such conditions, high cognitive sophistication can indeed facilitate more elaborate rationalizations and skillful defense of prior commitments. Acknowledging these limits strengthens the theoretical contribution of MS2R: the framework does not claim that reasoning is always biased but specifies when and why deliberation shifts from error correction toward identity defense.

## 6. Countermeasures and prospects

The mechanisms reviewed above suggest that MS2R is not an inescapable flaw of analytic thinking, but a context-sensitive pattern that emerges when identity or expressive goals dominate accuracy goals and when tasks permit selective interpretation. Therefore, countermeasures need to address both motivation (identity threat, reputational incentives) and cognition (how evidence is processed), as well as the broader information environment. The following subsections group candidate interventions into motivational, cognitive/analytical, and misinformation-focused strategies, with attention to their empirical support, limitations, and relevance for health psychology.

### 6.1. Facing reality: What individuals and institutions can do

One family of interventions seeks to reduce the identity threat that triggers defensive reasoning, thereby allowing System 2 to operate in a more accuracy-oriented mode. Self-affirmation interventions ask individuals to reflect on core personal values or domains of competence before encountering threatening information. In experimental studies, brief value-affirmation tasks have reduced derogation of dissonant sources and increased openness to counter-attitudinal evidence, including in health-risk contexts where participants confronted information about risky behaviors they engaged in.<sup>61,63</sup> Meta-analytic evidence in the health domain indicates that self-affirmation can yield small-to-moderate improvements in message acceptance and health behavior, although effects vary substantially by population, behavior, and implementation.<sup>67</sup> There are also null or attenuated effects when identity stakes are extreme or when affirmations are poorly matched to the targeted threat, underscoring that self-affirmation is helpful but not a panacea.

Identity-safe and value-concordant framing provides a complementary strategy. Rather than directly challenging core group identities, messages are framed to resonate with

the audience's existing values. For example, environmental policies can be framed as expressions of patriotism and system-preserving stewardship, which increases acceptance among system-justifying or conservative audiences.<sup>68</sup> In health communication, pro-vaccination or public health messages can emphasize values, such as protecting vulnerable family members, safeguarding community freedom, or fulfilling moral duties, depending on the audience's moral priorities.<sup>8</sup> These approaches are theoretically grounded in IPC: by reducing the perceived threat to identity, they lower the need for System 2 to engage in defensive rationalization. Empirically, value-concordant framing tends to reduce immediate resistance and sometimes narrows partisan gaps, but effects can be modest and context-dependent, and may not generalize across issues or groups.<sup>8,68</sup>

Therefore, both self-affirmation and identity-safe framing target MS2R at the motivational level. They aim to ensure that when System 2 is engaged, its goal is not to defend a threatened identity but to evaluate evidence under conditions of relative psychological safety. Nevertheless, their mixed effectiveness suggests that motivational interventions are most promising when combined with structural changes to incentives and reasoning tasks, rather than being treated as stand-alone solutions.

### 6.2 Cognitive and analytical interventions: Strengthening accuracy goals and symmetric scrutiny

A second family of interventions focuses on how System 2 processes information, attempting to heighten accuracy goals and enforce symmetric scrutiny of congenial and uncongenial evidence.

Accuracy prompts and incentives: A growing body of work shows that simple accuracy prompts—asking people to rate the accuracy of a headline before sharing it—or small monetary incentives for correct answers reliably improve truth discernment and reduce belief in false information, including COVID-19 misinformation.<sup>9,33,45</sup> Accuracy prompts and bonuses often shrink partisan gaps in factual beliefs, suggesting that a substantial portion of apparent motivated reasoning reflects under-engagement of analytic processing rather than irredeemable bias.<sup>9,10,34,51</sup> However, such incentives rarely eliminate partisan differences altogether, particularly on deeply entrenched identity-laden beliefs, indicating that MS2R can persist even when accuracy is salient. From a theoretical perspective, these findings support the idea that System 2's function is conditional: strengthening accuracy motivation reorients analytic processing toward error detection but does not fully override identity-related goals.

Symmetric-scrutiny protocols: “Consider the opposite” instructions and steel-manning exercises explicitly require individuals to generate the strongest arguments and evidence for the opposing position, or to list what evidence would change their minds. Classic experimental work shows that asking participants to consider reasons their initial judgments might be wrong reduces biased assimilation and leads to more moderate and balanced evaluations.<sup>41</sup> In deliberative or educational settings, structured debate formats and adversarial collaboration can institutionalize symmetric scrutiny by assigning roles and rewarding high-quality engagement with counterarguments.<sup>11</sup> These methods directly target confirmation bias and motivated skepticism, pushing System 2 to interrogate congenial claims as critically as uncongenial ones. At the same time, they impose cognitive load and are effortful; effects are often smaller when participants are fatigued, disinterested, or highly identified with their position, and they are challenging to scale outside structured environments.

Probabilistic reasoning and forecasting training: Training individuals to express beliefs as probabilistic forecasts and to track their calibration over time offers another route to channel analytic thinking toward accuracy. In the Good Judgment Project, participants who received structured training, feedback, and opportunities for practice substantially outperformed controls in geopolitical forecasting, with gains maintained over time.<sup>69</sup> Techniques such as decomposing problems into base rates and case-specific information, aggregating diverse judgments, and providing regular feedback systematically improved forecast accuracy and reduced overconfidence. In principle, similar methods could be applied to health-related beliefs (e.g., forecasting infection risk under different behaviors, or the impact of policy interventions). Evidence from forecasting research suggests that such training is most effective among motivated participants and requires sustained engagement; short one-off exercises deliver smaller benefits. Nonetheless, these methods directly counter MS2R by making error detection an explicit performance goal, backed by measurable feedback.

Task and information design: Finally, interventions can redesign tasks and information displays to reduce ambiguity and constrain opportunities for selective interpretation. Clear, transparent data visualizations, neutral labeling, and preregistered decision criteria for accepting or rejecting evidence all reduce the degrees of freedom available for MS2R to operate. Resource-rational models, such as BIASR, imply that various confirmation-like patterns arise when people update jointly on content and source reliability under uncertainty.<sup>29,48</sup> Clarifying what counts as good evidence in advance, separating

judgments of claim credibility from judgments of source credibility, and presenting statistical information in user-friendly formats (e.g., natural frequencies, icon arrays) can mitigate the rationalization of identity-congruent interpretations while remaining superficially “reasonable.” Empirical evaluations of such design changes are still emerging, but initial work in risk communication and misinformation correction suggests that structured, transparent presentations improve comprehension and reduce misinterpretation, especially among numerate individuals.<sup>8,46</sup>

Overall, cognitive and analytical interventions seek to steer System 2 away from advocacy and toward auditing by elevating accuracy goals, enforcing balanced reasoning, and limiting opportunities for selective scrutiny. Their effects are generally reliable but modest in size and can be attenuated under strong identity involvement, underscoring the importance of combining them with motivational and environmental strategies.

### 6.3. Misinformation-focused and information-environment interventions

A third set of strategies directly targets the information ecology within which MS2R operates, particularly regarding health misinformation.

Prebunking (psychological inoculation): Inoculation theory proposes that exposing individuals to weakened forms of misleading arguments, along with pre-emptive refutations and explanations of manipulative techniques, builds resistance to future misinformation. Experimental work on climate change has shown that short inoculation messages warning about misinformation tactics (e.g., “fake experts”) can preserve acceptance of scientific consensus even after exposure to misleading counter-messages.<sup>70</sup> Recent research has extended this approach using interactive games and videos that teach common manipulation strategies (e.g., emotional language, scapegoating), demonstrating improved resilience to misinformation across topics, including political and health-related content.<sup>35–38</sup> These interventions appear to work partly by shifting attention to argumentative quality and by pre-activating skepticism toward manipulative cues, thereby influencing how System 2 interprets later messages.

In health psychology, prebunking has been applied to COVID-19 misinformation and vaccine myths, often with encouraging but not uniform results. Inoculation messages typically reduce belief in targeted misinformation and can increase intentions to follow public health recommendations, but effect sizes are moderate and may wane over time without booster interventions.<sup>8,36</sup> Moreover,

prebunking tends to be more effective when messages are timely, concrete, and tailored to prevalent misinformation themes, suggesting that ongoing monitoring and updating are required.

**Debunking and corrective information:** Post-hoc corrections—fact-checks, myth-versus-fact formats, and explanatory refutations—remain central tools. Meta-analytic evidence indicates that corrections usually reduce belief in false claims, including health-related misinformation, although backfire effects appear less common than once feared.<sup>8</sup> However, corrections often leave residual misbeliefs, and repetition of myths can inadvertently reinforce familiarity-based “truthiness,” especially among individuals weak in analytic thinking or strongly motivated to protect identity. Therefore, debunking interacts with MS2R: when identity stakes are high, corrections may be scrutinized more critically or dismissed as biased, limiting their impact.

**Structural and platform-level changes:** Beyond individual messages, structural interventions—such as algorithmic downranking of repeatedly debunked content, labels that highlight contested claims, and promotion of high-quality health information sources—can alter the balance of evidence individuals encounter. Although empirical evaluation is still developing, such measures can be interpreted through resource-rational and selective-trust lenses: by making reliable health information more salient and misleading content less visible or clearly flagged, they change the inputs on which System 2 operates.<sup>30,46</sup> These interventions tackle MS2R indirectly, by reshaping the environment that channels IPC and expressive responding.

Across misinformation-focused interventions, evidence suggests that pre-emptive strategies (prebunking) often outperform purely reactive corrections, and that both need to be integrated with motivational and cognitive supports to achieve durable effects. Their particular relevance for health psychology lies in their ability to protect health beliefs before misinformation becomes entwined with identity and partisan conflict.

#### 6.4. Synthesis and implications for health psychology

Taken together, existing evidence indicates that no single intervention reliably neutralizes MS2R. Motivational interventions can reduce identity threat and reduce the likelihood of defensive reasoning, but their effects are heterogeneous and sometimes transient.<sup>61,63,67</sup> Cognitive and analytical interventions—accuracy prompts, symmetric scrutiny, probabilistic training, and clear task design—robustly improve accuracy and reduce generic susceptibility to misinformation, yet often leave residual

identity-congruent gaps.<sup>9,33,51,69</sup> Misinformation-focused strategies, such as prebunking and platform-level changes, mitigate exposure and build resistance, but their effects may fade without reinforcement and can be circumvented by novel misinformation tactics.<sup>8,36,37</sup>

For health psychology, the most promising approach appears to be multi-pronged and context sensitive. In settings where health beliefs are strongly identity-linked—such as vaccination, reproductive health, or nutrition norms embedded in cultural or religious identities—effective intervention likely requires: (i) reducing identity threat and aligning messages with core values (self-affirmation, identity-safe framing); (ii) heightening accuracy goals and structuring deliberation (accuracy prompts, symmetric scrutiny, forecasting and feedback); and (iii) shaping the information environment to minimize exposure to uncorrected misinformation and to inoculate audiences in advance (prebunking, reliable-source promotion). Educational programs for health professionals and the public can integrate these elements by teaching analytic skills and critical-thinking habits, while simultaneously cultivating identity-safe spaces and inoculation against common manipulation tactics.

At the research level, intervention studies that adopt preregistered, multi-arm designs and measure both directional bias and overall accuracy will be especially informative. Such studies can disentangle when interventions reduce identity-congruent distortion, when they simply raise accuracy uniformly, and when they mainly affect expressive responding. Process-tracing methods (e.g., response times, reasoning protocols), psychophysiological measures, and long-term follow-ups can clarify whether interventions genuinely shift underlying reasoning goals or primarily alter surface-level responses. In this sense, MS2R is best viewed not as an immutable property of System 2, but as a design problem for health communication and education—one that can be mitigated, though not fully solved, by carefully combining motivational, cognitive, and environmental levers.

## 7. Conclusion

This review has synthesized research from dual-process theory, motivated reasoning, social identity, political psychology, neuroscience, and resource-rational modeling to offer a coherent account of MS2R. Rather than treating System 2 as a uniformly de-biasing auditor, the analysis portrays deliberative reasoning as conditional in function: under accuracy-oriented goals, System 2 can correct intuitive errors, but when identity, expressive, or reputational motives dominate, the same analytic machinery can entrench bias by selectively scrutinizing and rationalizing information.<sup>3–5,7</sup> By explicitly incorporating

identity involvement, expressive responding, and selective trust into dual-process theory, the review clarifies why greater cognitive sophistication sometimes predicts polarization rather than consensus and delineates the boundary conditions under which this occurs. Applying these insights to health-relevant attitudes and behaviors—such as vaccine uptake, risk perception, and adherence to public health guidance—extends MS2R theory into a domain of direct interest to health psychology.<sup>8,33,46</sup>

Theoretically, the discussion refines classical dual-process models in two main approaches. First, it disentangles ability, disposition, and state engagement: numeracy and cognitive reflection are not interchangeable with momentary System-2 engagement, and high ability does not guarantee that deliberation will be deployed for accuracy rather than advocacy.<sup>47,57,58</sup> Second, it integrates multiple mechanisms—confirmation bias and motivated skepticism, cognitive dissonance and conflict rationalization, IPC, expressive responding and instrumental rationality, argumentative theory, and bounded/resource-rational approaches—into a layered framework that explains both robust directional motivated reasoning and the mixed evidence for the interactions between ability and identity.<sup>6,11–13,29,48</sup> High-powered replications and re-analyses suggest that identity-congruent responding is widespread. However, “the smartest are most biased” is not a general rule; rather, identity salience, task ambiguity, incentives, and audience visibility appear to be decisive in determining whether System 2 acts as an auditor or a defense attorney.<sup>19–22</sup>

Practically, this refined perspective carries important implications for health psychology and health communication. It cautions against interventions that rely solely on “more facts” or generic critical-thinking training under the assumption that additional information and numeracy will automatically depolarize contentious health issues, such as vaccinations, climate-related health risks, or pandemic control measures.<sup>8,10,23</sup> When health beliefs are tightly bound to social identities, moral values, or community norms, deliberation can be quickly co-opted to defend the status quo, discount dissonant evidence, and question the credibility of out-group sources.<sup>31,46</sup> Therefore, we underscore that effective health communication must consider identity, incentives, and information ecology, not just cognitive skill. Interventions, such as self-affirmation, identity-safe and value-concordant framing, accuracy prompts, symmetric-scrutiny protocols, probabilistic forecasting with feedback, clarity-enhancing information design, and prebunking, can shift the balance of motives so that System 2 is more likely to serve error detection than identity defense.<sup>8,9,37,61,67,69,70</sup>

At the same time, the evidence reviewed here points to clear gaps and priorities for future research. One priority is to causally disentangle the motivational and cognitive drivers of MS2R. Future experiments should systematically manipulate identity salience, accuracy incentives, and task ambiguity in factorial designs, while measuring ability and reflective disposition separately (e.g., numeracy, CRT, open-minded thinking). Such studies can clarify when directionally motivated reasoning is primarily driven by identity-protective motives, expressive signaling, or resource-rational selective trust, and when higher ability simply enhances accuracy regardless of identity.<sup>9,10,48</sup> A second priority is process-tracing and psychophysiological work that pinpoints when, in the reasoning sequence, bias enters: eye-tracking, response-time analysis, think-aloud protocols, and neural or autonomic markers of conflict and affect can help differentiate affect-first gating, dissonance-driven rationalization, and purely instrumental expressive responding.<sup>12,13,32</sup>

A third avenue lies in cross-domain and cross-cultural research, especially within health contexts. Most MS2R evidence still centers on United States partisan politics and climate change, leaving open questions about generalizability to other societies, health systems, and issue domains, such as chronic disease management, diet, reproductive health, and mental health stigma. Comparative work across cultures, political systems, and digital media environments could test whether the same mechanisms and boundary conditions observed in political domains apply to health-related beliefs and behaviors, and how variations in trust in institutions, media environments, and cultural values moderate MS2R.<sup>30,46</sup> Longitudinal and developmental studies could further examine whether susceptibility to motivated reasoning is a stable, trait-like tendency or whether it can be durably shaped by education, media literacy training, and repeated exposure to structured reasoning tasks.<sup>33,37,69</sup>

Finally, there is a need for rigorous intervention science grounded in the MS2R framework. Preregistered, multi-arm trials that compare motivational (e.g., self-affirmation, identity-safe framing), cognitive (e.g., accuracy prompts, symmetric scrutiny, forecasting), and ecological (e.g., prebunking, platform-level changes) approaches—alone and in combination—would clarify which levers are most effective, for whom, and under what conditions.<sup>8,36,38</sup> Such studies should measure both directional bias (identity-congruent versus incongruent errors) and overall accuracy, as well as expressive responding, so that reductions in polarization can be distinguished from genuine improvements in calibrated belief. Open-science practices, multi-lab collaborations, and attention to real-world

outcomes (e.g., vaccination rates, adherence to medical advice) will be crucial to ensure that theoretical insights about MS2R translate into meaningful advances in health behavior and communication.

In conclusion, understanding when and why System 2 reasoning shifts from correcting to defending bias is essential in an era of polarization and pervasive misinformation. The evidence reviewed here suggests that more thinking is not always better thinking when identities, values, and reputational stakes are on the line; analytic reasoning can become a powerful ally of motivated cognition. Recognizing this dual potential of System 2 does not undermine the value of critical thinking, but rather calls for designing contexts—in education, health communication, and public discourse—in which accuracy goals are aligned with identity and social incentives. For health psychology in particular, integrating the MS2R perspective promises more realistic theories of belief formation and change, and more effective strategies to foster open-minded, evidence-based reasoning about health in diverse, identity-rich societies.

## Acknowledgments

None.

## Funding

This study is supported by Faculty of Education, Palacký University under an institutional research grant ‘Selected topics of psycho-logical research in the field of educational psychology’ (IGA\_2025\_013).

## Conflict of interest

The authors declare no competing interests.

## Author contributions

*Conceptualization:* Hongyang Liu, Jana Kvintová  
*Methodology:* Hongyang Liu, Jana Kvintová, Irena Plevová  
*Writing – original draft:* Hongyang Liu  
*Writing – review & editing:* Jana Kvintová, Lucie Váchová, Irena Plevová, Justyna Dockalova

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data

Not applicable.

## Further disclosure

The paper has been uploaded to or deposited in a preprint server ([https://doi.org/10.31234/osf.io/y254w\\_v2](https://doi.org/10.31234/osf.io/y254w_v2)). The authors utilized ChatGPT 5.0 for language polishing during manuscript preparation. AI-edited sections underwent human review and proofreading to ensure content accuracy. The authors assume responsibility for this process. Since none of the authors are native English speakers, this article was translated using ChatGPT 5.0.

## References

1. Kahneman D. Fast and slow thinking. *Allen Lane and Penguin Books, New York*. 2011;
2. Stanovich KE, West RF. Advancing the rationality debate. *Behavioral and Brain Sciences*. 2000;23(5):701-717.  
doi: 10.1017/S0140525X00623439
3. Kunda Z. The case for motivated reasoning. *Psychological Bulletin*. 1990;108(3):480-498.  
doi: 10.1037/0033-2909.108.3.480
4. Taber CS, Lodge M. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*. 2006/07/01 2006;50(3):755-769.  
doi: 10.1111/j.1540-5907.2006.00214.x
5. Kahan DM, Peters E, Wittlin M, et al. The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*. 2012/10/01 2012;2(10):732-735.  
doi: 10.1038/nclimate1547
6. Kahan DM. Misinformation and identity-protective cognition. *Yale Law & Economics Research Paper*. 2017;(587)
7. Kahan DM, Peters E, Dawson EC, Slovic P. Motivated numeracy and enlightened self-government. *Behavioural Public Policy*. 2017;1(1):54-86.  
doi: 10.1017/bpp.2016.2
8. Ecker UKH, Lewandowsky S, Cook J, et al. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*. 2022/01/01 2022;1(1):13-29.  
doi: 10.1038/s44159-021-00006-y
9. Bago B, Rand DG, Pennycook G. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *J Exp Psychol Gen*. Aug 2020;149(8):1608-1613.  
doi: 10.1037/xge0000729
10. Pennycook G, Rand DG. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. 2019/07/01/ 2019;188:39-50.

- p>doi: 10.1016/j.cognition.2018.06.011
11. Mercier H, Sperber D. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*. 2011;34(2):57-74.  
doi: 10.1017/S0140525X10000968
  12. van Veen V, Krug MK, Schooler JW, Carter CS. Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*. 2009/11/01 2009;12(11):1469-1474.  
doi: 10.1038/nn.2413
  13. Westen D, Blagov PS, Harenski K, Kilts C, Hamann S. Neural bases of motivated reasoning: an fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election. *J Cogn Neurosci*. Nov 2006;18(11):1947-58.  
doi: 10.1162/jocn.2006.18.11.1947
  14. Cullen CM, Sanders HM, Chung KC. It Is Time to Rethink Our Approach to Bias in Medicine. *Plastic and Reconstructive Surgery*. 2025;155(5):745-752.  
doi: 10.1097/PRS.00000000000011915
  15. Tay SW, Ryan PM, Ryan CA. Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. *Canadian Medical Education Journal*. 10/18 2016;7(2):e97-103.  
doi: 10.36834/cmej.36777
  16. Damjanović K, Novković V, Pavlović I, Ilić S, Pantelić S. A Cue for Rational Reasoning: Introducing a Reference Point in Cognitive Reflection Tasks. *Europe's Journal of Psychology*. 02/28 2019;15(1):25-40.  
doi: 10.5964/ejop.v15i1.1701
  17. Sunstein CR. Hazardous Heuristics. *The University of Chicago Law Review*. 2003;70(2):751-782.  
doi: 10.2307/1600596
  18. Simis MJ, Madden H, Caciatore MA, Yeo SK. The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science*. 2016/05/01 2016;25(4):400-414.  
doi: 10.1177/0963662516629749
  19. Persson E, Andersson D, Koppel L, Västfjäll D, Tinghög G. A preregistered replication of motivated numeracy. *Cognition*. 2021/09/01/ 2021;214:104768.  
doi: 10.1016/j.cognition.2021.104768
  20. Stagnaro MN, Tappin BM, Rand DG. No association between numerical ability and politically motivated reasoning in a large US probability sample. *Proceedings of the National Academy of Sciences*. 2023/08/08 2023;120(32):e2301491120.  
doi: 10.1073/pnas.2301491120
  21. Strömbäck C, Andersson D, Västfjäll D, Tinghög G. Motivated reasoning, fast and slow. *Behavioural Public Policy*. 2024;8(3):617-632.
  22. Tappin BM, Pennycook G, Rand DG. Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General*. 2021;150(6):1095-1114.  
doi: 10.1037/xge0000974
  23. Cho J, Cokely ET, Ramasubramanian M, Allan JN, Feltz A, Garcia-Retamero R. Numeracy does not polarize climate change judgments: Numerate people are more knowledgeable and knowledge is power. *Decision*. 2024;11(2):320-344.  
doi: 10.1037/dec0000223
  24. Lodge M, Taber CS. *The rationalizing voter*. Cambridge University Press; 2013.
  25. Schaffner BF, Luks S. Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us about the source of Political Misinformation in Surveys. *Public Opinion Quarterly*. 2018;82(1):135-147.  
doi: 10.1093/poq/nfx042
  26. Festinger, L., 1954. *A theory of social comparison processes*. *Human Relations*, 7(2), 117-140.  
doi: 10.1177/001872675400700202
  27. Harmon-Jones E, Mills J.(2019) Cognitive dissonance: Reexamining a pivotal theory in psychology, 2nd ed. <https://psycnet.apa.org/doi/10.1037/0000135-000>
  28. Nickerson RS. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*. 1998/06/01 1998;2(2):175-220.  
doi: 10.1037/1089-2680.2.2.175
  29. Lieder F, Griffiths TL. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*. 2020;43:e1. e1.  
doi: 10.1017/S0140525X1900061X
  30. O'Connor C, Weatherall JO. *The misinformation age: How false beliefs spread*. Yale University Press; 2022.
  31. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. *Annual Review of Public Health*. 2020;41 (Volume 41, 2020):433-451.  
doi: 10.1146/annurev-publhealth-040119-094127
  32. Lois G, Tsakas E, Yuen K, Riedl A. Tracking politically motivated reasoning in the brain: the role of mentalizing, value-encoding, and error detection networks. *Social Cognitive and Affective Neuroscience*. 2024;19(1):nsae056.  
doi: 10.1093/scan/nsae056
  33. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. *Nature*. 2021/04/01

- 2021;592(7855):590-595.  
doi: 10.1038/s41586-021-03344-2
34. Pennycook G, Rand DG. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*. 2022/04/28 2022;13(1):2333.  
doi: 10.1038/s41467-022-30073-5
  35. Biddlestone M, Roozenbeek J, Suiter J, Culloty E, van der Linden S. Tune in to the prebunking network! Development and validation of six inoculation videos that prebunk manipulation tactics and logical fallacies in misinformation. *Political Psychology*. 2025/03/27 2025; 46, 1858 1886.  
doi: 10.1111/pops.70015
  36. Maertens R, Roozenbeek J, Simons JS, et al. Psychological booster shots targeting memory increase long-term resistance against misinformation. *Nature Communications*. 2025/03/11 2025;16(1):2062.  
doi: 10.1038/s41467-025-57205-x
  37. Roozenbeek J, van der Linden S, Goldberg B, Rathje S, Lewandowsky S. Psychological inoculation improves resilience against misinformation on social media. *Science Advances*. 2022;8(34):eabo6254.  
doi: 10.1126/sciadv.abo6254
  38. van der Linden S, Kyrychenko Y. A broader view of misinformation reveals potential for intervention. *Science*. 2024/05/31 2024;384(6699):959-960.  
doi: 10.1126/science.adp9117
  39. Lord CG, Ross L, Lepper MR. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*. 1979;37(11):2098-2109.  
doi: 10.1037/0022-3514.37.11.2098
  40. Kraft PW, Lodge M, Taber CS. Why People “Don’t Trust the Evidence”: Motivated Reasoning and Scientific Beliefs. *The ANNALS of the American Academy of Political and Social Science*. 2015/03/01 2015;658(1):121-133.  
doi: 10.1177/0002716214554758
  41. Lord CG, Lepper MR, Preston E. Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*. 1984;47(6):1231-1243.  
doi: 10.1037/0022-3514.47.6.1231
  42. Bonica A. Mapping the Ideological Marketplace. *American Journal of Political Science*. 2014/04/01 2014;58(2):367-386.  
doi: 10.1111/ajps.12062
  43. Mason LH. Partisan sorting and behavioral polarization in the American electorate. 2012: <https://ssrn.com/abstract=2105031>
  44. Spry A. *Identity in American Politics: A Multidimensional Approach to Study and Measurement*. Columbia University; 2018.  
doi: 10.7916/D8C26CWM
  45. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*. 2020/07/01 2020;31(7):770-780.  
doi: 10.1177/0956797620939054
  46. Beck M, Ahmed R, Douglas H, et al. Motivated Reasoning and Risk Governance: What Risk Scholars and Practitioners Need to Know. In: Gattinger M, ed. *Democratizing Risk Governance: Bridging Science, Expertise, Deliberation and Public Values*. Springer International Publishing; 2023:29-53.
  47. Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*. Oct 2011, 2024-12-03 2011;39(7):1275-1289.  
doi: 10.3758/s13421-011-0104-1
  48. Pilgrim C, Sanborn A, Malthouse E, Hills TT. Confirmation bias emerges from an approximation to Bayesian reasoning. *Cognition*. 2024/04/01/ 2024;245:105693.  
doi: 10.1016/j.cognition.2023.105693
  49. Akerlof GA, Kranton RE. Economics and identity. *The quarterly journal of economics*. 2000;115(3):715-753.
  50. Akerlof GA, Kranton RE. Identity economics: How our identities shape our work, wages, and well-being. *Identity Economics*. Princeton University Press; 2010.  
doi: 10.1515/9781400834181
  51. Prior M, Sood G, Khanna K. You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*. 2015;10(4):489-518.
  52. Lenz GS. *Follow the leader?: how voters respond to politicians’ policies and performance*. University of Chicago Press; 2012.
  53. Mayhew DR. *Congress: The electoral connection*. Yale university press; 2004.
  54. Lois G, Tsakas E, Riedl A. Facts over partisanship: Evidence-based updating of trust in partisan sources. *Journal of Experimental Psychology: General*. Published online August 21, 2025.  
doi: 10.1037/xge0001815
  55. Henderson L, Gebharter A. The role of source reliability in belief polarisation. *Synthese*. 2021/12/01 2021;199(3):10253-10276.  
doi: 10.1007/s11229-021-03244-y
  56. von Sydow M, Merdes C, Hahn U. The Temporal Dynamics

- of Belief-based Updating of Epistemic Trust: Light at the End of the Tunnel? *arXiv preprint arXiv:191213380*. 2019;
57. Evans JSBT, Stanovich KE. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*. 2013/05/01 2013;8(3):223-241.  
doi: 10.1177/1745691612460685
58. Frederick S. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. 2005;19(4):25-42.  
doi: 10.1257/089533005775196732
59. Lodge M, Taber CS. The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis. *Political Psychology*. 2005/06/01 2005;26(3):455-482.  
doi: 10.1111/j.1467-9221.2005.00426.x
60. Slovic P, Finucane M, Peters E, MacGregor DG. Rational actors or rational fools: implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*. 2002/01/01/ 2002;31(4):329-342.  
doi: 10.1016/S1053-5357(02)00174-9
61. Cohen GL, Aronson J, Steele CM. When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin*. 2000/11/01 2000;26(9):1151-1164.  
doi: 10.1177/01461672002611011
62. Fein S, Spencer SJ. Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*. 1997;73(1):31-44.  
doi: 10.1037/0022-3514.73.1.31
63. Sherman DK, Cohen GL. The Psychology of Self-defense: Self-Affirmation Theory. *Advances in Experimental Social Psychology*. Academic Press; 2006:183-242.
64. Downs A. An Economic Theory of Democracy. *Harper and Row*. 1957 1957;28  
doi: 10.1086/257897
65. Kuran T. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press; 1998.
66. Noelle-Neumann E. The Spiral of Silence a Theory of Public Opinion. *Journal of Communication*. 1974;24(2):43-51.  
doi: 10.1111/j.1460-2466.1974.tb00367.x
67. Epton T, Harris PR, Kane R, van Koningsbruggen GM, Sheeran P. The impact of self-affirmation on health-behavior change: A meta-analysis. *Health Psychology*. 2015;34(3):187-196.  
doi: 10.1037/hea0000116
68. Feygina I, Jost JT, Goldsmith RE. System Justification, the Denial of Global Warming, and the Possibility of "System-Sanctioned Change". *Personality and Social Psychology Bulletin*. 2010/03/01 2009;36(3):326-338.  
doi: 10.1177/0146167209351435
69. Mellers B, Ungar L, Baron J, et al. Psychological Strategies for Winning a Geopolitical Forecasting Tournament. *Psychological Science*. 2014/05/01 2014;25(5):1106-1115.  
doi: 10.1177/0956797614524255
70. van der Linden S, Leiserowitz A, Rosenthal S, Maibach E. Inoculating the Public against Misinformation about Climate Change. *Global Challenges*. 2017/02/01 2017;1(2):1600008.  
doi: 10.1002/gch2.201600008