

Bypassing as a Technique for Refuting Disinformation: Contributions from Neuroscience and Behavioral Economics and Impacts on Law

José Jance Marques Grangeiro*

University of Brasília (Universidade de Brasília – UnB), Brasília, Brazil

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100600952>

Received: 22 June 2026; Accepted: 27 June 2026; Published: 09 July 2026

ABSTRACT

Over the last two decades, research on countering disinformation has consolidated a confrontational refutation paradigm grounded in the direct factual correction of false information. Recent work in social psychology and behavioral science has proposed, as an alternative or complement, a non-confrontational strategy known as bypassing, which seeks to attenuate the impact of disinformation by introducing beliefs whose evaluative implications are opposite to those of the misinformation, without ever mentioning the false content. Objectives: this article examines the bypassing technique and assesses its potential to inform the legal design of disinformation regulation in Brazil. Methods: it conducts a theoretical and interdisciplinary review that articulates three bodies of literature — the neuroscience of cognitive defensiveness, the behavioral economics of dual-process reasoning and nudges, and Brazilian and comparative legal scholarship on platform regulation and freedom of expression. Results: the review shows that factual correction, although effective on average, is constrained by the continued influence effect, by the affective discomfort that retraction provokes, and by identity-protective cognition, whereas bypassing lowers the affective cost of attitude change by leaving the original belief untouched and supplying a salient correlate belief at the moment of judgment. Conclusions: acknowledging the limits of factual correction and the side effects of direct confrontation invites legal scholarship to rethink regulatory instruments, shifting part of the normative effort from content suppression toward fostering healthier informational environments. Bypassing, articulated with systemic risk-mitigation duties and transparency mechanisms, offers a path compatible with the constitutional protection of freedom of expression, although it requires careful calibration to prevent corporate capture and political instrumentalization.

Keywords: bypassing; disinformation; neuroscience; behavioral economics; legal regulation.

INTRODUCTION

Disinformation has become a central object of the regulatory agenda of contemporary democracies, owing to the scale at which it circulates on digital platforms and the reach of its effects on public deliberation [1]. The phenomenon cuts across sensitive domains such as electoral integrity, public health, scientific debate, the formation of legal opinion, and the protection of vulnerable groups, heightening the urgency of institutional responses. For at least a decade, the specialized literature has documented paradigmatic cases such as interference in elections, the circulation of anti-vaccine content during the Covid-19 pandemic, and the spread of denialist discourse around well-established historical facts [13][12].

The dominant institutional response has combined three main fronts: professional fact-checking, the labeling of false content by the platforms, and the enactment of mandatory-moderation statutes, such as the passage of Bill 2630 in Brazil and the Digital Services Act in the European Union [1]. The three fronts share the assumption that the best response to false information is its identification and direct refutation before the user, through removal, labeling, or explicit counter-speech. The underlying paradigm is confrontational: it operates by acknowledging falsity, exposing the contradiction, and expecting the recipient to revise the belief.

Empirical research on debunking, however, has produced evidence that factual correction, although effective on average, operates with relevant limitations [4]. Disinformation continues to influence judgments after it has been retracted, a phenomenon described as the continued influence effect, partly because of the cognitive discomfort generated by the retraction itself [10][15]. When the recipient has a strong ideological attachment to the false content, direct confrontation tends to produce reactance, defensiveness and, under certain conditions, reinforcement of the original belief [6].

It is in this scenario that the concept of bypassing, formulated by Granados Samayoa and Albarracín [8] as a non-confrontational strategy for attitude change, gains prominence. The proposal consists in shifting the focus of the intervention away from the false belief and toward a correlate belief with an opposite evaluative implication, capable of reorienting the recipient's attitude without the disinformation being mentioned at all. The technique challenges the academic common sense that had been treating factual correction as the privileged instrument, while at the same time offering a new repertoire for the design of public policy and regulatory obligations.

This article examines bypassing along three interdependent dimensions. The first is the neurocognitive substrate of resistance to correction, with particular attention to the brain mechanisms that make confrontation a path of high affective cost. The second is the analytical toolkit of behavioral economics concerning dual processes of cognition, within which bypassing conceptually approximates accuracy-prompt techniques and the nudge program. The third is the legal reading of the technique, aimed at identifying how regulatory instruments can incorporate non-confrontational principles without compromising the constitutional guarantees of freedom of expression.

The rationale for the study stems from an observable mismatch between behavioral evidence and normative design. Brazilian legislation on disinformation has so far privileged mechanisms of removal and platform liability, with little attention to behavioral alternatives already validated in international research [1]. The dialogue between law and behavioral science, although not unprecedented, still lacks operational translation for the specific theme of disinformation.

The text is organized into four central chapters and a concluding section. The second section presents the conceptual foundations of bypassing, in contrast with the confrontational paradigm of debunking. The third examines the contribution of neuroscience to understanding cognitive defensiveness and the advantage of lower-resistance pathways. The fourth articulates bypassing with the behavioral-economics literature on dual processes, accuracy prompts and nudges. The fifth unfolds the legal implications of the technique for the regulation of disinformation in Brazil, with attention to freedom of expression and platform liability.

From Correction to Circumvention: The Conceptual Foundations of Bypassing

The confrontational refutation paradigm starts from a simple causal chain: disinformation produces false beliefs, false beliefs generate deviant attitudes, and these attitudes guide socially undesirable behavior [10]. The typical intervention therefore consists in identifying the false information, refuting it with factual correction, and thereby rebalancing the recipient's chain of beliefs and attitudes [4]. The operationalization of this logic gave rise to a vast repertoire of techniques, ranging from textual correction to visual labeling, including the insertion of warnings and referral to verified sources.

The meta-analysis conducted by Chan et al. [4] gathered fifty-two experimental studies on debunking and confirmed that corrections produce medium-to-large reductions in the impact of disinformation, compared with control conditions. The aggregate effect is robust, but the distribution of results is heterogeneous, with strong sensitivity to the format of the correction, the profile of the recipient, and the type of content refuted. The subsequent literature went on to investigate, in greater detail, the factors that moderate the effectiveness of corrections, in particular their structure, their timing, and the psychological characteristics of the receiver [2][6].

This research converged on a set of best practices, now incorporated into science-communication and fact-checking manuals. Corrections are more effective when they make explicit the disinformation to be refuted, offer a coherent alternative explanation, reinforce the source of the counter-speech, and avoid the excessive repetition

of the false information [10]. There are also specific recommendations about the timing of the correction, with it generally being preferable to correct before or immediately after the initial exposure [2].

Despite the methodological consolidation of debunking, the literature records persistent limitations of factual correction. Lewandowsky et al. [10] describe the continued influence effect as the tendency of disinformation to keep guiding inferences even after it has been explicitly retracted and acknowledged as false by the recipient. The phenomenon has been replicated across multiple contexts, from autobiographical memory to the formation of political attitudes, and challenges the naive hypothesis that mere exposure to the truth suffices to neutralize falsity.

Susmann and Wegener [15] demonstrate experimentally that part of the continued influence effect derives from the discomfort produced by the retraction itself, and not only from limitations of memory or processing. When a person is confronted with the information that something they believed is not true, an increase in negative affective responses is observed that lowers the probability of belief revision, even when the evidence presented is technically convincing. The finding has significant practical implications, because it suggests that the affective cost of correction may cancel part of its cognitive benefit.

There is also evidence that corrections may produce the so-called backfire effect in specific subpopulations, although the generality of this effect is debated in the more recent literature [6]. The central issue is not so much the frequency with which the original belief is reinforced after correction, but rather the finding that, to varying degrees, direct confrontation generates defensiveness, reactance and psychological costs that limit its effectiveness. DeVerna et al. [5], for example, recorded an ideological asymmetry in the failure to correct disinformation, with recipients aligned with the original content revising their beliefs to a lesser degree.

Bypassing, as set out by Granados Samayoa and Albarracín [8], inverts part of the logic of debunking: instead of attacking the false belief, an alternative belief with an opposite evaluative implication is introduced, capable of shifting the resulting attitude without the original disinformation being mentioned. The formulation draws on classic models of social psychology, especially the expectancy-value theories of Fishbein and Ajzen [7], which conceptualize attitude as a weighted sum of beliefs about an object. If attitude results from the integration of multiple beliefs, then it is possible to alter it by introducing new beliefs, without having to refute the pre-existing ones.

In an experiment conducted by Calabrese and Albarracín [3], participants exposed to a false news report about the supposed carcinogenic effects of genetically modified foods had their regulatory attitudes modified both by correction and by a message highlighting the contribution of such foods to the preservation of bees. The two interventions produced statistically equivalent effects on attitudes toward restrictive policies and on intentions to support such policies. The experimental design was replicated across three independent studies, with the same direction of results, providing robust evidence for the equivalence.

In a later study, using headlines in a format analogous to the everyday consumption of news on social media, bypassing proved more effective than correction when recipients formed judgments from memory, and equivalent when the judgment was formed simultaneously with exposure [8]. The comparative advantage of bypassing appears to relate to the fact that it supplies, at the moment of decision, more salient information on which the attitude is built, as opposed to correction, which requires the active retrieval of the debunked content. The technique, therefore, does not replace debunking, but offers a complementary route for situations in which confrontation tends to be unproductive.

Framing bypassing as a general strategy also makes it possible to group under the same heading phenomena previously treated in separate literatures. Granados Samayoa and Albarracín [8] draw attention to the parallel between bypassing applied to disinformation and the literature on counter-stereotype training, which shows reductions in the activation of automatic stereotypes when participants affirm counter-exemplars instead of negating stereotypical associations. The same principle appears in studies on the lateral generalization of attitudes, in which persuasion about a correlate object produces change in the target object, without direct confrontation.

Bypassing In Light of Neuroscience: Cognitive Defensiveness and the Path of Least Resistance

The persistence of disinformation after correction can be partly explained by relatively well-mapped neurocognitive mechanisms. Ecker et al. [6] identify three classes of factors that make disinformation resistant to correction: the familiarity generated by repeated exposure, the narrative coherence between the false information and the individual's prior repertoire, and the identity bond between the belief and the recipient's self-image. Each of these factors operates through distinct brain processes, but they converge in the same direction: factual correction, by making explicit the contradiction between belief and reality, activates circuits associated with symbolic threat, with increased affective reactivity and a narrowing of the disposition for cognitive revision.

Familiarity plays a central role in processing fluency, described by the literature as the subjective sense of ease with which information is understood [10]. Familiar information is processed with less effort and tends to be evaluated as more truthful, by virtue of what has come to be called the truth-by-repetition effect. The repetition of disinformation during the very act of correction may, paradoxically, reinforce the sense of veracity of the refuted content, requiring redoubled care in the design of the counter-speech.

Narrative coherence, in turn, depends on the embedding of disinformation in an explanatory schema that lends meaning to multiple correlate events. When this coherence is high, removing an isolated piece of the narrative leaves intact the broader schema, which continues to operate as an interpretive guide [6]. The literature on conspiracy theories demonstrates this phenomenon with particular clarity, by showing that the refutation of a specific element tends to be absorbed as one more indication of the cover-up alleged by the conspiratorial narrative itself.

Susmann and Wegener [15] advanced the understanding of the continued influence effect by demonstrating that part of the phenomenon derives from the discomfort produced by retraction, and not only from working-memory limitations. In controlled experiments, manipulating the discomfort associated with correction altered the magnitude of the effect, suggesting an affective component that adds to the cognitive components described by earlier literature. The finding dialogues with decades of research on cognitive dissonance, which describe the human tendency to avoid information that contradicts previously assumed identity commitments.

Affective neuroscience offers a basis for this observation by showing that contradiction with core beliefs activates brain regions associated with social pain, similar to those activated in situations of exclusion and of threat to self-image [6]. The reaction is not merely cognitive, but affective-bodily, with effects on attentional regulation and on the disposition to sustain careful deliberation. The practical consequence is that factual correction, when directed at identity-laden beliefs, tends to recruit defensive responses that reduce its effectiveness.

Beyond discomfort, there are moderating factors linked to individual differences in the propensity for analytical reflection, in engagement in motivated reasoning, and in trust in specific sources [13]. More reflective people tend to benefit more from corrections, whereas recipients with low analytical engagement or strong partisan alignment with the false content may remain impervious. The heterogeneity of effects therefore demands a segmented approach, in which different techniques are triggered according to the cognitive profile of the audience.

Bypassing reduces the affective cost of the change process by leaving the recipient's original belief untouched. By introducing an alternative belief with an opposite evaluative implication, the intervention circumvents the activation of identity-defense mechanisms, allowing the attitude to reorganize without the subject having to acknowledge publicly, or even internally, the falsity of the initial information [8]. The strategy acts on the evaluative outcome of the set of beliefs, without disputing each element of that set point by point.

This property is especially important in contexts of high polarization, in which the defense of the false belief functions as a marker of group belonging. The literature on identity-protective cognition, developed in dialogue with the theory of cognitive dissonance, shows that correction tends to be interpreted as an attack on the group, not as neutral information [6]. Under such conditions, confrontation reinforces the internal cohesion of the polarized group and amplifies resistance to revision.

The non-confrontation operated by bypassing, by offering an alternative cognitive route, mobilizes mental resources at a lower affective cost. Instead of demanding from the recipient a labor of identity deconstruction, the technique asks only for the integration of a new belief into the cognitive repertoire, with effects on the resulting global attitude [8]. The neuroscience of resistance to change thus provides an explanatory basis for a phenomenon that experimental psychology had already been documenting: paths of least resistance tend to produce more behavioral change than paths of direct confrontation.

Behavioral Economics and Bypassing: From Dual Processes to The Architecture of Informational Choices

Over recent decades, behavioral economics has provided a consolidated vocabulary for describing the limits of human rationality in the face of ambiguous, noisy or abundant information. Kahneman, Sibony and Sunstein [9] systematize two sets of decision errors: biases, which produce systematic deviations in the direction of a response, and noise, which produces random variability among judgments that should be equivalent. The circulation of disinformation operates on both, by exploiting heuristic shortcuts that reduce the cognitive cost of processing large volumes of digital content.

The dual-process model, consolidated by Kahneman and by successive generations of researchers, distinguishes between a fast, automatic and associative system and a slow, deliberate and analytical one [9]. Most consumption of content on social media occurs under the dominance of the first system, with low attention to veracity, a heavy weighting of social cues such as the number of likes and shares, and strong sensitivity to emotional appeals. This configuration explains why false content with a high emotional charge tends to circulate more quickly than corrections that are technically precise but affectively neutral.

Pennycook and Rand [13] argue, on the basis of broad empirical evidence, that a significant part of the sharing of disinformation does not stem from a sincere conviction as to the veracity of the content, but from mere inattention. In several experiments, the same participants who correctly classified a headline as false, when asked to assess its accuracy, declared a willingness to share it when the task was framed as a sharing decision. The finding opens space for interventions that act not on belief, but on the user's attention at the moment of the decision.

Pennycook and Rand [13], in a meta-analysis of twenty experiments and about twenty-six thousand participants, demonstrate that accuracy prompts — which merely draw the user's attention to the importance of considering the veracity of content before sharing it — reduce the sharing of false news by around ten percent, without significantly affecting the sharing of true news. The effect does not stem from a change of belief, but from the momentary redirection of attentional focus, in line with the hypothesis that a significant part of disinformation sharing is driven by inattention rather than by conviction. The intervention operates on the architecture of the decision, without having to refute any content.

The research extended to different prompt formats, from questioning the importance of sharing only accurate content to inviting the user to evaluate a neutral headline before the main task [13]. The effects were replicated across different platforms, across different thematic contexts, including politics and public health, and in populations with different demographic profiles. The magnitude of the effect is modest but robust, and the instrument has the advantage of not requiring platforms to make any judgment about the substantive truth of the content in circulation.

The independence between the effectiveness of the prompt and the substance of the content is particularly valuable from a regulatory standpoint. Instead of transferring to the platform the burden of identifying what is true or false — which involves obvious risks of overblocking and ideological capture — the obligation concentrates on the design of the decision environment [13]. The literature further suggests that combining accuracy prompts with other behavioral interventions tends to produce additive effects, amplifying the aggregate impact on the quality of shared content.

There is an evident conceptual proximity among accuracy prompts, bypassing and the nudge program proposed by Sunstein [14]. In all three, behavioral change occurs through a contextual modification of the most salient cognitive pathway, and not through direct argumentative persuasion. The difference is that bypassing substitutes

a correlate belief for the target belief, whereas the accuracy prompt merely reactivates a pre-existing value of the user, and the classic nudge alters the choice environment without modifying the repertoire of available information [14][13].

The three techniques share the assumption that interventions on the intuitive thinking system tend to be more effective than deliberative interventions in low-attention contexts. This assumption underpins the program of libertarian paternalism, which defends the admissibility of interventions that steer choices without suppressing alternatives [14]. The normative justification of the nudge, however, depends on the intervention being transparent, reversible and calibrated to promote the recipient's well-being, rather than serving third-party interests.

Applying these categories to bypassing demands additional caution. The introduction of alternative beliefs, even true ones, may operate as manipulation if it occurs without the recipient's awareness or without the possibility of informed revision [14]. The behavioral literature, while validating the effectiveness of the technique, warns that its large-scale adoption must be accompanied by mechanisms of transparency about the criteria for selecting the correlate beliefs, on pain of turning the informational architecture into a vector of heteronomous influence over public opinion.

The discussion of the ethical admissibility of such interventions dialogues with the classic controversy over the boundary between legitimate persuasion and manipulation. Sunstein [14] argues that the transparency of the techniques, combined with the preservation of the recipient's freedom of choice, keeps the nudge within acceptable ethical limits, though he acknowledges the points of tension. Bypassing, by not revealing to the recipient that an intervention on their attitude is under way, sits in a more sensitive zone, requiring that its institutional incorporation be accompanied by specific accountability guarantees.

Legal Implications: Disinformation Regulation, Freedom of Expression and Platform Liability

The legal regulation of disinformation faces a structural dilemma. On the one hand, state omission in the face of the massive circulation of false content compromises constitutional goods such as the integrity of the electoral process, public health, and the dignity of vulnerable groups [1]. On the other hand, interventions centered on content removal and on the criminalization of expression tend to produce side effects on freedom of expression, especially in politically polarized contexts, in which the category of fake news is frequently appropriated by state actors to delegitimize legitimate criticism [11].

Pasquetto, Lim and Bradshaw [12] record that, since 2016, more than one hundred and twenty journalists have been arrested under anti-disinformation laws in various countries, frequently for reporting that ran counter to government interests. The figure is eloquent about the risk of instrumentalization of such normative frameworks by authoritarian or authoritarian-leaning governments. Regulatory discussion, therefore, cannot dispense with a careful assessment of the political economy of the proposed instruments, with particular attention to the possibility that they may be captured by actors opposed to the very goods they claim to protect.

Brazilian legal scholarship on freedom of expression in the digital age warns that poorly calibrated regulatory responses tend to favor actors with greater technical and financial capacity to litigate, displacing public debate beyond the reach of ordinary citizens [11]. The very expansion of private moderation mechanisms creates points of asymmetry, in which the decisions of opaque platforms affect different social sectors unequally. The design of the regulatory regime must therefore internalize these asymmetries and provide review mechanisms accessible to recipients without specialized litigation capacity.

The recognition, by the behavioral literature, that direct corrections produce limited and occasionally counterproductive effects offers grounds for a regulation less centered on suppression. The removal obligation, although intuitive, transfers to the platform the judgment about the substantive truth of the content, with risks of overblocking, excessive standardization, and the silencing of legitimate speech that approaches, without coinciding with, the frontier of disinformative content [1]. The European regulatory experience with the Digital Services Act shows that mandatory-removal instruments demand a high degree of institutional sophistication to avoid excesses.

Pasquetto, Lim and Bradshaw [12] argue that disinformation research should abandon the search for generalizable answers and invest in local understandings of when and how misleading content affects specific communities. This methodological shift has direct regulatory consequences, because it disqualifies normative designs that treat disinformation as a homogeneous category, demanding instead instruments calibrated by context, community and platform. Suppression-centered regulation tends, by its binary nature, to disregard these differences.

Empirical research on debunking further suggests that the late removal of disinformative content may be less effective than early labeling or than the simultaneous presentation of correlate content with an opposite evaluative implication [2][8]. The behavioral literature therefore converges with critical legal scholarship in pointing to the limits of repressive regulation and in inviting the exploration of preventive and architectural instruments.

Bypassing offers a model of intervention compatible with the constitutional guarantees of freedom of expression. Instead of obliging platforms to remove controversial content, with the known risks of overblocking, regulation may require that they expose the user to correlate content with an opposite evaluative implication, or that they insert accuracy prompts at critical points in the sharing flow [13][14]. The regulatory obligation falls on the design of the informational environment, and not on the substantive truth of the statements in circulation, which reduces the risk of the State becoming the definitive arbiter of what may be said.

Framing this obligation as a duty to mitigate systemic risk finds a parallel in the European Digital Services Act, which imposed on very large platforms duties to assess and reduce risks relating to electoral integrity, public health and the protection of minors [1]. The difference is that the focus shifts from the removal of individual items to the aggregate configuration of the informational ecosystem, with obligations of audit, transparency and accountability. Bypassing may integrate this design as one technique among others, complementary to correction, labeling and the algorithmic prioritization of verified quality sources.

The modulation of the obligation by platform size is legally relevant. Large-reach platforms, with the capacity to significantly influence the formation of public opinion, may be subject to more robust duties of exposure to correlate content, whereas smaller services, with negligible systemic effect, remain under less intrusive regimes. This graduation avoids the regulatory overcost that would strike small actors and concentrates the normative effort where the negative externality is greatest.

Introducing these instruments into the Brazilian legal order demands careful articulation with the Internet Civil Framework (Marco Civil da Internet) and with the case law of the Federal Supreme Court on provider liability. The obligation of exposure to alternative content is distinguished from the removal obligation, because it does not presuppose a judgment of unlawfulness, and it dialogues with the notion of a duty to mitigate systemic risk, present in the European debate [1]. The construction of this duty in Brazilian law may rest on constitutional foundations such as the social function of property, consumer protection and the protection of democracy, without implying a revision of the rule of subjective liability for third-party content.

The modulation of these obligations by platform size, by content category and by temporal context — especially during electoral periods and health emergencies — allows the regulatory design to remain proportionate to the legal goods protected, without turning private moderation into outsourced state censorship. The electoral period, by force of a specific constitutional provision, accommodates more intense regulatory duties, in particular as regards the integrity of the democratic process [11]. In more diffuse domains, such as cultural debate or the dispute over historical narratives, regulatory intensity should be lower, on pain of producing disciplining effects on the public sphere.

The dialogue of bypassing with the case law of the Federal Supreme Court on provider liability demands attention to recent precedents that discussed the constitutionality of Article 19 of the Internet Civil Framework. Although the specific design of the obligations stems from a legislative choice, the jurisprudential recognition of positive duties of mitigation creates room for incorporating non-confrontational behavioral instruments. Articulating these obligations with the data-protection regime established by the General Data Protection Law

(LGPD) requires additional caution as to the use of personal information for the personalization of prompts and correlate content, especially in sensitive contexts.

The adoption of bypassing techniques as an element of a positive regulatory obligation must therefore be accompanied by mechanisms of transparency about the criteria for selecting the alternative content, on pain of the informational architecture coming to operate as an instrument of corporate manipulation, now backed by law. Behavioral science, while offering the instrument, warns of the risks of its capture by heteronomous interests [14]. The regulatory obligation, to fulfill its purpose, must make explicit objective selection criteria, open space for contestation, and provide for independent auditing.

Critical scholarship on disinformation further warns of the risk that researchers and regulators may incorporate, unreflectively, epistemic assumptions proper to the position they occupy, turning the definition of disinformation into a vector of ideological bias [12]. The operationalization of bypassing ultimately depends on judgments about which correlate beliefs should be presented as a counterpoint, which involves sensitive choices. The adoption of public standards, with the participation of multiple voices, is a condition for these choices not to become ideological imposition disguised as technical neutrality.

There are also long-term risks associated with users' habituation to behavioral techniques. If the interventions operate by surprise or by low attention, it is possible that their effectiveness diminishes over time, as recipients learn to recognize and cognitively correct for them [14]. The literature on accuracy prompts has not yet converged on the pace of this decay, but the hypothesis suggests that regulation cannot treat these techniques as definitive solutions, but rather as components of a constantly updated repertoire.

Concluding Remarks

Bypassing constitutes a non-confrontational refutation technique whose effectiveness is sustained by robust experimental evidence and by a theoretical basis articulated in the cognitive neurosciences and in behavioral economics. Factual correction remains irreplaceable in domains in which the truth of the proposition matters in itself, as in science, professional journalism and expert testimony in court. For the purpose of mitigating the attitudinal and behavioral effects of disinformation in public debate, however, the literature suggests that circumventive strategies tend to be as effective as, or more effective than, direct confrontation [8].

Migrating this finding to the field of law demands caution. The temptation to convert behavioral evidence into a direct regulatory obligation runs up against the heterogeneity of the contexts in which disinformation circulates and the risks of corporate capture of the informational design [12]. The most promising path appears to consist of systemic-mitigation obligations, calibrated by risk and accompanied by transparency mechanisms, which replace, wherever possible, the paradigm of content suppression with the paradigm of a healthy architecture of choices.

The articulation proposed among neuroscience, behavioral economics and law does not purport to offer a finished solution to the problem of disinformation. Each of the three areas operates with its own epistemic assumptions, and their convergences do not dispense with the careful translation that legal design requires. The contribution of this study lies in pointing out that the behavioral literature today possesses instruments that exceed the dominant confrontational paradigm, and that their incorporation into law can broaden the regulatory repertoire without aggravating the risks to freedom of expression.

Future research should map, with greater precision, the conditions under which bypassing outperforms correction, the domains in which confrontation remains preferable, and how the Brazilian legal order can incorporate these distinctions without compromising the protection of freedom of expression. The research agenda should include empirical studies calibrated to the Brazilian context, with attention to the linguistic, cultural and institutional particularities that differentiate the national informational ecosystem from the contexts in which the international literature was predominantly developed. At this point, the encounter among neuroscience, behavioral economics and law ceases to be merely an interdisciplinary curiosity and becomes a condition for the regulation of disinformation to gain effectiveness without sacrificing constitutional principles.

Finally, it should be noted that the responsible adoption of these techniques depends on an institutional culture committed to transparency and accountability. The recent history of the use of behavioral technologies by governments and corporations offers sufficient examples to justify this caution. The challenge for law, in the present, is to incorporate the lessons of the behavioral sciences without reproducing the mistakes of those who treated their findings as neutral tools of social management.

REFERENCES

1. Barreto, A. M. P., Porto, C. S., & Jaborandy, C. C. M. (2024). A sociedade de desinformação e seus impactos no direito brasileiro. *Thoth*.
2. Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5), e2020043118.
3. Calabrese, C., & Albarracín, D. (2023). Bypassing misinformation without confrontation improves policy support as much as correcting it. *Scientific Reports*, 13, 6005.
4. Chan, M.-P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546.
5. DeVerna, M. R., Guess, A. M., Berinsky, A. J., Tucker, J. A., & Jost, J. T. (2024). Rumors in retweet: Ideological asymmetry in the failure to correct misinformation. *Personality and Social Psychology Bulletin*, 50(1), 3–17.
6. Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1, 13–29.
7. Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.
8. Granados Samayoa, J. A., & Albarracín, D. (2024). Bypassing as a non-confrontational influence strategy. *Current Opinion in Psychology*, 59, 101855.
9. Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.
10. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
11. Macedo Junior, R. P. (2018). Liberdade de expressão ou dever de falar a verdade? In M. Barbosa (Ed.), *Pós-verdade e fake news: Reflexões sobre a guerra de narrativas*. Cobogó.
12. Pasquetto, I. V., Lim, G., & Bradshaw, S. (2024). Misinformed about misinformation: On the polarizing discourse on misinformation and its consequences for the field. *Harvard Kennedy School Misinformation Review*, 5(5).
13. Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, 13, 2333.
14. Sunstein, C. R. (2014). *Why nudge? The politics of libertarian paternalism*. Yale University Press.
15. Susmann, M. W., & Wegener, D. T. (2022). The role of discomfort in the continued influence effect of misinformation. *Memory & Cognition*, 50, 435–448.