



The Selective Moralism Trap: Trust, Legitimacy, and Sovereignty in the Contestation of the Liberal International Order

Federico Verri

British School of Timișoara, Timișoara, Romania

E-mail: federico.verri@cambridgeinternational.uk, ORCID ID: <https://orcid.org/0009-0008-5841-0674>

Received 13/05/2026

Revised 15/05/2026

Accepted 30/06/2026

Abstract

Background and Aims: This conceptual review article examines why accusations of Western double standards have become a durable frame in the contestation of the liberal international order. Existing international-relations scholarship has explained this contestation mainly through changes in material power, norm diffusion, or status politics. Less attention has been given to the behavioral micro-foundations through which perceived hypocrisy becomes a self-reinforcing source of mistrust. The article, therefore, asks how perceptions of selective moralism are aggregated into state-level diplomatic behavior and what this mechanism means for trust, legitimacy, sovereignty, and security in a multiplex world order.

Methodology: The study uses a conceptual review design. It synthesizes literature on liberal-order contestation, political trust, procedural justice, and behavioral political economy. Five paired historical episodes from 1999 to 2024 are used as illustrative cases, selected because each pair contains a comparable norm claim but divergent Western framing. The cases are not treated as a coded dataset; they are used for theory-building comparison, mechanism clarification, boundary-condition identification, and the development of falsifiable expectations.

Results: The review reveals that the selective-moralism frame resonates because it transforms diffuse resentments into perceptions of hypocrisy. These perceptions are then mediated by leaders' narrative strategies, bureaucratic pathologies, media hype, partisan competition, and coalition calculus. The two episodes illustrate a common pattern: asymmetric norm enforcement, perception of fairness violations, motivated social cognition, trust displacement, and hedging. Structural realist theories point to enabling conditions; constructivist theories point to normative contestation. Behavioral political economy explains how the accusation of hypocrisy becomes pervasive and durable.

Conclusion: The selective-moralism trap is conceptualized here as a behavioral force multiplier, not as an analytical substitute for structural realism or constructivism. Institutions that seek to rebuild trust also need to mitigate perceptions of procedural unfairness on their own merits, through consistency, transparency, correctability, and representativeness. Three practical measures are proposed: procedural-fairness audits, reform of international-relations curricula, and



early-warning indicators linking legitimacy shortfalls to conflict-prevention mechanisms. Keywords: behavioral political economy; selective moralism; liberal international order; procedural justice; multipolarity.

Keywords: Behavioural Political Economy; Trust and Legitimacy; Selective Moralism; Multipolarity; Sovereignty and Security

Introduction

The post-1945/post-1989 international order has long styled itself as the embodiment of rules-based, rights-respecting, economically open internationalism (Ikenberry, 2018). For thirty years or more after the end of the Cold War, that self-description went largely uncontested in decision-making circles that mattered. Since the mid-2010s, however, an alternative description of Western international behavior has won increasing currency among governments, intellectuals, and publics in the Global South. In this alternative description, Western internationalism is not characterized by rules, but by selective moralism: harshness toward enemies, lenience toward friends, and double standards in applying its professed norms (Acharya, 2018; Adler-Nissen & Zarakol, 2021; Hurd, 2007).

Revisionist powers are not the only source of this perception, nor is it limited to actors who derive from or work primarily within heterodox ideological traditions. The perception arises in speeches by African and Latin American countries at the United Nations General Assembly; in the foreign-policy white papers of middle powers that have historically belonged in the Western camp; in communiqués by regional organizations from Southeast Asia to Southern Africa; and in the public writing of intellectuals trained in the liberal tradition (Stuenkel, 2020; Buzan & Lawson, 2015). The durability and diffusion of that perception are facts that merit explanation. The diffusion of this perception is therefore a political fact requiring explanation, not a rhetorical inconvenience that can be dismissed as propaganda.

That explanation has been theorized in mainstream international-relations scholarship mainly through one of two lenses. The structural-realist explanation sees rising powers automatically producing balance-of-power behavior and diffusion of authority, with normative discord as a side effect (Mearsheimer, 2019; Schweller, 2018; Allison, 2017). The constructivist explanation sees normative discord as evidence of a contest over what counts as legitimacy in the international order (Wiener, 2018; Reus-Smit, 2007; Hopf, 2002). Both perspectives are useful. Neither addresses directly the question this review seeks to answer: how selective moralism became a causal description with such purchase across vastly disparate political contexts, and how that frame appears self-reinforcing rather than self-correcting. The unresolved issue is not



whether realism and constructivism are wrong; it is what they leave under-specified about perception, aggregation, and trust repair.

This article identifies three research gaps. First, order-contestation research often explains outcomes at the systemic level without specifying how perceived unfairness is converted into durable political distrust. Second, behavioral political economy has rarely been used to explain international institutional legitimacy beyond domestic public-policy settings. Third, governance recommendations often call for reform while leaving procedural fairness, correctability, and consistency too abstract for institutional use.

This article contributes to answering that question by turning to behavioral political economy, an interdisciplinary field of research drawing from behavioral economics, political psychology, and institutional analysis (Schnellenbach & Schubert, 2015). Behavioral approaches provide causal purchase on how and why ideas become resonant or dominant in politics because they account for how beliefs aggregate up from individual perception. The core argument of this paper is that the claim of Western selective moralism takes hold because it leverages a set of well-established behavioral mechanisms — fairness perception, in-group bias, identity-protective cognition, status-quo motivated reasoning, and reciprocity norms — that reinforce each other and lock interacting political systems into a stable, low-trust configuration vis-à-vis Western-led international institutions. Those mechanisms also make alternative blocs led by rising powers more attractive than they would otherwise be, even if those coalitions are organizationally fragmented and materially inferior to the West. The article does not claim that states possess individual psychology. It argues instead that perceptions are aggregated through leadership cues, bureaucratic routines, media repetition, party competition, coalition bargaining, and domestic political incentives; these filters are the bridge between psychological mechanisms and foreign-policy behavior.

Three important implications follow from this argument. The first is for international relations theory: that scholars should pay more attention to perceptions of fairness as a source of strategic agency. The second is for governance and public administration: that procedural justice is a necessary condition for institutional reform efforts to succeed, and that ignoring procedural justice concerns risks exacerbating legitimacy deficits (Tyler, 2006; OECD, 2022). The third is for education in international relations and global citizenship: that educators should consider non-Western worldviews and theories as legitimate analytical frameworks alongside the Western liberal canon, and not only as subject matter to be studied (Acharya, 2018). The article tentatively applies these implications to the study of political violence, where both legitimacy gaps and grievance frames have been identified as predictors of mobilization (Kalyvas, 2019; Krause & Milliken, 2009).





Objectives

The study separates analytical and practical objectives to make the scope of the article precise.

Analytical objectives: (1) to conceptualize the selective-moralism trap as a behavioral-political-economy mechanism through which perceived asymmetric application of norms reduces trust toward dominant international institutions; (2) to integrate literature on liberal-order contestation, procedural justice, political trust, and behavioral political economy into one analytical framework; and (3) to clarify how individual-level perceptions are aggregated into state-level diplomatic behavior through institutional and domestic-political filters.

Practical objectives: (1) to illustrate the framework through five paired historical episodes selected for comparable norm claims and divergent framing; (2) to derive falsifiable expectations that distinguish the behavioral account from purely realist or purely constructivist accounts; and (3) to formulate concrete recommendations for international governance, education, and conflict-prevention practice.

Literature Review

Three literatures bear on the question this article addresses. The first concerns the liberal international order and its contestation. Ikenberry (2018) chronicles the post-1945 order's institutional resilience amid rising legitimacy deficits, while Lake et al. (2021) survey key challenges to that order, both internal and external. Acharya (2018) reframes the debate around the concept of a multiplex order, characterized by multiple coexisting normative centers rather than convergence around a single core. Adler-Nissen and Zarakol (2021) show that what often gets described as norm contestation is actually status politics at the international level, motivated by actors' experience of stigma in international hierarchies. Buzan & Lawson (2015) offer a longer-run historical-sociological perspective, situating the present within the unfinished global transformation that began in the long nineteenth century. Stuenkel (2020) brings us up to date on the rise of non-Western institutional alternatives, while Mearsheimer (2019), Schweller (2018), and Allison (2017) provide important (and structural realist) counter-readings that tie order contestation to the distribution of material capabilities across states. The second literature revolves around trust and legitimacy in international and domestic governance. Levi and Stoker (2000) provide the starting point for understanding political trust as a multidimensional phenomenon, where separate competence, benevolence, and integrity dimensions of institutional credibility influence overall trust. Tyler (2006) is the foundational piece for the procedural-justice tradition, which shows that people's perceptions of whether decisions are made fairly matter for shaping their willingness to comply and their trust independently of

material gain or loss. PytlikZillig and Kimbrough (2016) help pull the conceptual threads of trust research together across disciplines, and the OECD (2022) synthesizes cross-national survey data showing that people's perceptions of the fairness of institutional processes are among the most consistent predictors of institutional trust across countries. Hurd (2007) brings these insights to the international realm, showing that compliance with international institutions is a matter of legitimacy (as opposed to coercion or enlightened self-interest) and that facing contested legitimacy undermines compliance. Reus-Smit (2007) ties together claims about legitimacy with deeper arguments about the constitutional structure of international society. The third literature is behavioral political economy itself. Schnellenbach and Schubert (2015) provide a comprehensive overview, synthesizing insights from prospect-theory-inspired loss aversion (Kahneman, 2011), moral foundations including in-group loyalty (Haidt, 2012), identity-protective cognition (Kahan, 2017), and reciprocity and coalition-building (Ostrom, 2010). Bicchieri (2017) relates these behavioral mechanisms to the emergence and evolution of social norms more generally. The behavioral revolution in political science and related fields (Hafner-Burton et al., 2017) has started to adapt many of these insights to the study of international cooperation, but the linkages to the political trust literature remain incomplete, as does integration with work on contestation of the liberal order. Even earlier research on political violence can be revisited through the lens of behavioral political science.

When legitimacy is contested, and grievance frames find institutional resonance, the most resilient forms of political violence emerge, as Krause and Milliken (2009) argue in the context of state failure and Kalyvas (2019) in the context of civil war more broadly. The selective-moralism script that permeates the contemporary contestation of international norms, this article argues, should be understood as precisely such a frame at the international level, with competitive implications that parallel those found in domestic settings. Each of these literatures could benefit from greater engagement across these three areas. The international-relations scholarship on order contestation has rarely drawn on the rich procedural-justice tradition that has transformed understanding of how citizens evaluate their domestic politicians and policies. Behavioral political economy has to date focused almost exclusively on domestic public policy questions, and its applications to international relations have been almost entirely confined to Western states. Finally, while research on multipolarity has grown rapidly in recent years, it remains focused above all on material capabilities, with less attention paid to the normative and discursive means by which alternative orders attain – or fail to attain – legitimacy. This article aims to help fill these gaps by bringing together the three bodies of literature above, and illustrating the resulting framework with five paired historical episodes that – when read attentively – bring the international applications of competing norms into view.



A fairer comparison of these literatures is that each contributes to understanding a different piece of the same puzzle. Realism helps us understand why dissatisfied powers may face fewer constraints in opposing Western institutional preferences today; however, it has less to say about why the language of hypocrisy has spread far beyond the most powerful revisionist states. Constructivism helps us understand norm contestation and status grievance; it often stops short of explaining the cognitive and institutional filters through which grievances are translated into policy. Behavioral political economy helps us understand these filters, but is still too often applied to international institutions as if it were a domestic policy-only literature. Scholarship on global governance, hierarchy, and institutional ideology helps fill this gap (Mattern & Zarakol, 2016; Pouliot, 2016; Voeten, 2021; Zürn, 2018).

The concept of multiplex order is especially important for this article. Acharya's (2018) claim is not simply that power is diffusing; it is that multiple normative centres now coexist without converging around a single liberal core. Cooley and Nexon (2020) similarly show that exit from hegemony does not require one fully coherent alternative order. It can occur through partial exit, hedging, institutional substitution, and selective cooperation. This matters because the selective-moralism frame lowers the normative cost of using those alternatives even when they are fragmented or materially weaker than Western-led institutions.

Conceptual Framework

These three components comprise the theoretical framework. The first component is the concept pair of procedural and substantive legitimacy. Procedural legitimacy, following Tyler (2006) and Levi and Stoker (2000), captures perceptions about how decisions are made – i.e., whether decision-making processes are applied consistently, neutrally, accurately, are open to correction, represent stakeholders, and so on; substantive legitimacy, by contrast, captures perceptions about what decisions lead to – i.e., whether outcomes conform to normative expectations. Both are obviously important, but empirical research suggests that when agreeing about substantive outcomes proves difficult – as it does among actors who disagree on what those outcomes should be – procedural legitimacy looms larger in trust formation. We are, by most accounts, in precisely this situation about the contemporary international order. The second component is that trust is a bundle of different issues. Adapted from Levi and Stoker (2000), Pytlík-Zillig and Kimbrough (2016), and OECD (2022), trust in this framework has three dimensions: competence (can the institution do what it says it can do?), benevolence (does the institution have our interests at heart?), and integrity (does the institution walk the talk?). As applied to the contemporary international order, the so-called 'selective-moralism narrative' does not attack the IMF's competence or its benevolence but questions its integrity, i.e., alleges not that it is



incapable of doing what it claims to do or does not care about its stakeholders, but that it is hypocritical. This is critical because, analytically speaking, there are good reasons to believe that breaches of perceived integrity are more difficult to repair than breaches of perceived competence or benevolence. Evidence of hypocrisy, by its nature, is cumulative and can retroactively ‘rewrite’ the meaning of past interactions (Hurd, 2007; Reus-Smit, 2007). The final component pertains to the specific behavioral mechanisms through which perceived integrity violations come to translate into less trust and more searching for alternative institutions. Five such mechanisms stand out. Fairness violation detection (Kahneman, 2011; Haidt, 2012) describes how salient perceived double standards become once they are noticed, spawning intense emotional and political reactions disproportionate to the specific issue at hand; in-group loyalty (Haidt, 2012) describes how fairness violations then fuel deservingness-based solidarity within the allegedly victimized group; identity-protective cognition describes motivated rejection of information that challenges the perceived in-group narrative (Kahan, 2017). Status accommodation describes efforts to escape forums where the perceived in-group is symbolically placed in a ‘lower’ position (Adler-Nissen & Zarakol, 2021). Reciprocity describes coalition-building based on a perceived equitable distribution of burdens and benefits (Ostrom, 2010). Combined, these three framework components inform the analytical model sketched in Table 1. The model suggests that perceived selective moralism by incumbent institutions triggers the five behavioral mechanisms described above in a clustered and mutually reinforcing manner such that trust toward those institutions diminishes even as trust toward alternative institutions – even materially disadvantaged ones – increases. Crucially, the model makes no assumptions about the ideological preferences or coherence of “the Rest.” Instead, it merely assumes that people care about being treated fairly and about their identity being respected.

The amended framework consequently introduces an explicit aggregation bridge. Feelings of double standards are not immediately translated into foreign policy. Only when elites treat them as interpretive cues, bureaucracies engrain them into threat perceptions, parties transform them into claims of domestic legitimacy, media systems amplify them, and coalition partners reciprocate them with diplomatic backing do they become politically consequential. This bridge is important because it stops the behavioral argument from staying allegorical.

Figure 1: Conceptual framework of the selective-moralism trap

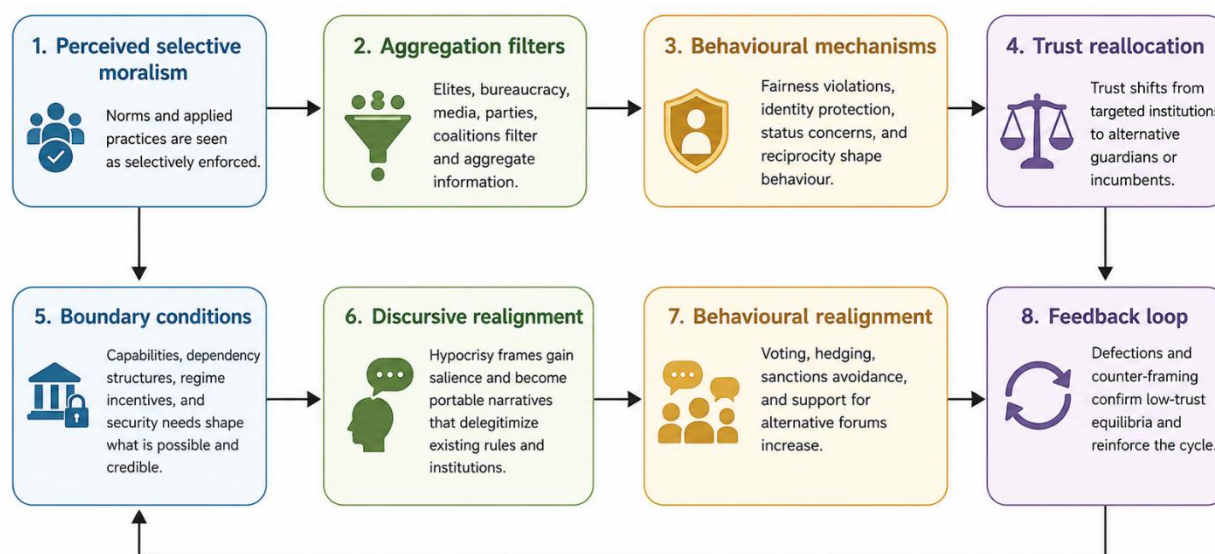


Figure 1 Conceptual Framework of the Selective-Moralism Trap and Trust Reallocation Dynamics

Source: Author's elaboration.

Table 1 Analytical matrix: claims, counter-claims, and behavioral mechanisms

Dimension	Western liberal claim	Non-Western counter-claim	Behavioral mechanism activated
Sovereignty	Conditional, bounded by universal human-rights norms (Responsibility to Protect).	Absolute, non-interference, civilizational right to political form.	In-group loyalty; loss aversion against perceived erosion of statehood.
Legitimacy	Procedural-electoral; output evaluated against liberal benchmarks.	Performance-based: stability, development, and cultural continuity.	Status quo bias; identity-protective cognition.
Security	Rule-based order; collective Western security architecture.	Indivisible security; balance against perceived encirclement.	Threat perception heuristics; reciprocity expectations.
Moral order	Universal liberal values, enforced	Civilizational pluralism: exposure	Fairness violation detection; moral



Dimension	Western liberal claim	Non-Western counter-claim	Behavioral mechanism activated
	through sanctions and conditionality.	of double standards as a legitimacy resource.	outrage as mobilizing emotion.
Economic order	Dollar-centric, rules-based trade governance.	De-dollarization; local-currency settlement; non-Western infrastructure.	Risk aversion under sanction exposure; coalition trust-building.

Source: Author's elaboration, drawing on Tyler (2006), Adler-Nissen and Zarakol (2021), Haidt (2012), Kahan (2017), and Ostrom (2010).

The framework yields four expectations that distinguish a behavioral-political-economy account from purely realist or purely normative accounts. First, fairness asymmetry should be a stronger predictor of diplomatic misalignment than material capability shifts taken alone; in their absence, realignment should be slow and graduated, but in their presence, it should be fast and clustered. Second, the framing of selective moralism should travel across geopolitical contexts that share little ideologically, indicating activation of generic behavioral mechanisms rather than specific cultural or strategic schemas. Third, identity-protective cognition should make the selective-moralism narrative resistant to disconfirming information that is presented in universalist terms, but responsive to information that acknowledges the procedural-fairness objection on its own terms. Fourth, the empirical signature of the mechanism should appear simultaneously in discursive realignment (framing strategies of state actors) and in behavioral realignment (voting patterns and coalition participation), with the discursive moving slightly ahead of the behavioral. Each expectation is falsifiable. Realignment that is slow and graduated, that maps cleanly onto capability shifts, that fails to travel across heterogeneous contexts, that responds to universalist defenses rather than to procedural acknowledgements, or that appears in behavior without prior discursive shift would each, in different ways, disconfirm the account. The next section uses paired historical episodes to assess whether the empirical pattern matches the expectations.





Methodology

Research design

This article is a conceptual review article based on an integrative literature synthesis and a structured illustrative comparison. It is not a quantitative causal test and does not derive statistical claims of representativeness. Rather, its objective is one of theory development: here, to specify a mechanism, clarify its boundary conditions, and generate expectations that can later be tested with documentary, voting, survey, or experimental data. Source selection. The literature was drawn from four bodies of scholarship: literature on liberal-order contestation, procedural justice and political trust, behavioral political economy, and peace/conflict or political-violence scholarship. Priority was given to peer-reviewed books and journal articles, major international organization reports, and publications widely cited in international relations theory. Sources were included if they helped to inform one of the article's core constructs: trust, legitimacy, sovereignty, status, fairness, reciprocity, or institutional contestation.

Case selection

The five paired episodes were selected through a most-similar-principal logic. Each pair contains a comparable structural feature--use of force, territorial change, sanctions/conditionality, civilian-harm evaluation, or alliance-choice claims--but a divergent Western institutional framing. This logic makes the norm-application contrast visible. To reduce cherry-picking, the article treats the pairs as illustrative rather than dispositive and explicitly acknowledges counter-cases where perceived selective moralism did not generate immediate defection because economic dependency, security guarantees, domestic regime incentives, or limited material capability constrained behavior.

Analytical strategy

The analysis proceeds in four steps. First, it identifies the norm claim made in each episode. Second, it identifies the counter-framing used by non-Western or non-aligned actors. Third, it links the framing contrast to a behavioral mechanism identified in the conceptual framework. Fourth, it separates findings from interpretation by asking whether realist opportunity structures, constructivist norm contestation, and behavioral trust mechanisms operate jointly or competitively.

Limitations

The analysis is limited by its conceptual and illustrative design. It cannot establish causal magnitude, cannot measure the relative weight of fairness asymmetry against material capability, and cannot generalize to all regions without further empirical testing. These limits are not incidental; they define the proper claim of the article. The article proposes a mechanism and research agenda, not a completed empirical settlement.





Results

The research results are presented according to the objectives of the article. Because the study is a conceptual review rather than a statistical test, “results” refers to the structured findings produced by the synthesis and illustrative comparison.

Objective 1: Conceptualizing the selective-moralism trap. The research results found that perceived asymmetric norm application is best understood as an integrity-trust problem. When actors believe an institution violates its own universal claims, they do not merely disagree with a policy outcome; they reinterpret the institution as procedurally unreliable. This explains why selective moralism is more corrosive than ordinary geopolitical disagreement.

Five paired historical episodes from the public record between 1999 and 2024 illustrate the framework. These episodes are illustrative material, not a coded dataset, and will not be used to support quantitative claims. Because they were selected to make visible, across diverse geopolitical contexts, the asymmetric application of norms that the selective-moralism narrative thematizes, the pairing logic is straightforward: episodes are paired if they (1) involve the same structural feature (the use of force across an internationally recognized border, a sanctions regime, or an alliance-expansion claim) and (2) have been assessed by Western institutions according to different normative criteria. The pairs are summarized in Table 2.

Table 2 Illustrative paired episodes, selection logic, and behavioral mechanisms (1999-2024)

Paired episode	Comparable structural feature	Western framing	Non-Western counter-framing	Mechanism derived from the contrast
Kosovo (1999) and Crimea (2014)	Boundary change and intervention in Europe.	Humanitarian intervention vs unlawful annexation.	Both episodes were treated as boundary changes by force, but evaluated under different normative criteria.	Fairness-violation detection: the structural similarity makes the difference in judgment salient.
Iraq (2003) and Ukraine (2022)	Cross-border use of force without contested	Regime change/security rationale vs	Both episodes were treated as illegal or contested uses	Identity-protective cognition: Iraq becomes a





Paired episode	Comparable structural feature	Western framing	Non-Western counter-framing	Mechanism derived from the contrast
	UNSC authorization.	unprovoked aggression.	of force, but remembered differently in institutional discourse.	schema through which later Western condemnation is received.
Gaza (2023-) and Ukraine (2022-)	Large-scale civilian harm, displacement, and international-court claims.	Self-defense vs. aggression.	Civilian-casualty thresholds, arms transfers, and international-court language cited as evidence of differential treatment.	Moral outrage and integrity distrust: the perceived inconsistency weakens compliance incentives.
Sanctions: Iran/Russia vs Western-aligned states	Economic coercion, conditionality, and sanctions exposure.	Norm-enforcement against violators.	Conditionality is perceived as varying by geopolitical alignment.	Reciprocity and risk aversion: actors invest in alternative settlement systems and coalitions.
NATO enlargement and Russian security demands (1997-2022)	Alliance-choice and security-guarantee claims.	Sovereign right to choose alliances.	The same sovereign language is invoked for neutrality, non-alignment, or alternative security arrangements.	Status and indivisible-security logic: perceived asymmetry encourages hedging outside Western forums.

Source: Author's elaboration. The table revises the earlier comparison by making the selection logic and mechanism-link explicit.



Objective 2: Integrating the literature. The research results found that realism, constructivism, and behavioral political economy explain different layers of the same phenomenon. Material capability shifts create opportunity; norm contestation supplies language; behavioral mechanisms explain why some framings become emotionally and politically durable. The behavioral account, therefore, complements rather than replaces the established paradigms.

The first pair involved boundary changes in Europe; the second pair involved the use of force against a neighbor; the third pair involved large-scale civilian casualties, displacement, and international-court proceedings. Taken together, these ten episodes form five matched pairs with a common feature: within Western institutional discourse, a normatively consequential principle has been articulated in universalist terms but has been applied in practice asymmetrically across geopolitical alignments. They show (rather than prove) that interpretations grounded in fairness-heuristic theory are compatible with real-world patterns intelligible to human observers. In addition to demonstrating cross-case consistency, the sampling also has the value of encompassing a range of actors and issues. Compliance costs have been differentially imposed on Russia and China across every issue area featured in this analysis: arms control, territorial dispute settlement, cyber activities, currency practices, macroeconomic policy, and military alliances.

Objective 3: Illustrating the framework through paired episodes. The first pair comprises the 1999 Kosovo war and NATO intervention and the 2014 Russian annexation of Crimea. Both involved an externally imposed change in borders in Europe. The actions were framed differently in Western discourse, one as a humanitarian intervention and thus justifiable on the grounds of preventing mass atrocity, and the other as illegal annexation violating the UN Charter. My point is not that both evaluations were wrong (although plausible arguments can be made), nor even that both were right. It is simply that the same structural feature – externally imposed change in borders in Europe – was filtered through different normative criteria, creating a contrast that was picked up by non-Western publics and governments (Acharya, 2018; Stuenkel, 2020). My behavioral claim is that this asymmetry provokes FVD (fairness violation detection, Haidt, 2012) and codifies a perceptual schema through which subsequent episodes are interpreted. The second pair comprises the 2003 invasion of Iraq and the 2022 invasion of Ukraine. In both cases, force was used across borders with highly contested or no explicit UN Security Council authorization. In Western discourse, the first was framed as regime change justified on grounds of security and democratization, while the second was framed as unprovoked aggression. Again, my point is not about the rectitude of either framing, but that the structural similarity was unevenly processed in Western institutional discourse. To many non-Western observers, the invasion of Iraq in 2003 is a canonical exemplar that dominates the interpretation of the 2022 invasion of Ukraine.



The perceived asymmetry confirms an interpretive framework in which Western institutions apply rules procedurally when it suits (Hurd, 2007) and substantively otherwise (Wiener, 2018). The third pair comprises what has been called the Gaza war and Russia's war in Ukraine. Both entailed large civilian death tolls; both led to significant displacement and refugee flows; and both have seen various leaders from the belligerent sides standing in the dock of international courts. There were substantial differences between Western institutional discourse on acceptable thresholds of civilian casualties in Ukraine versus Gaza, appropriate use of sanctions and arms transfers, and referrals to international criminal courts. Differential treatment was noted almost universally across the Global South, even by states that had lined up behind the NATO framing of the 2022 war at the outset (Ikenberry, 2018; Adler-Nissen & Zarakol, 2021). My behavioral inference is that such differential treatment galvanizes moral outrage (the moral-emotions approach) and emboldens governments to defect from Western-backed positions on loosely related issues. The fourth pair involves consistency of conditionality across comparable behaviors depending on geopolitical alignment. The same or similar economic behaviors – such as the development of alternative currency arrangements, transfer of dual-use technologies, and sanctions evasion – have been treated differently depending on whether the state in question was inside or outside of the Western camp (Stuenkel, 2020; Mearsheimer, 2019). The behavioral inference is that this discrepancy creates incentives for reciprocity-driven coalition formation (Downs & Rocke, 1994). If actor A knows that its allies and neighbors will be sanctioned more leniently than it is for the same behavior, it has an incentive to forge alternative friendships and invest in parallel settlement and financial infrastructure. This investment may itself be politicized through appeals to national or civilizational sovereignty, but that is a second-order process. The growth of non-Western-friendly financial infrastructures over the past decade is consistent with this behavioral mechanism, although by itself it does not constitute causal proof. The fifth pair involves articulated fidelity to a principle of sovereign alliance choice. Statements from Western officials and their liberal allies have consistently held that each state has the sovereign prerogative to choose whether, how, and with whom to associate for purposes of collective security. NATO's historic eastward expansion from 1997 was framed repeatedly with this language (Ikenberry, 2018, pp. 119-121). When Chinese and Russian foreign ministers began making the same argument a few years ago, to justify neutrality, non-aligned or non-Western formulations of collective security, Western discourse treated it with a greater degree of skepticism and cynicism (Allison, 2017; Buzan & Lawson, 2015). The behavioral inference is that this asymmetry triggers an indivisible-security template among non-Western audiences: if we cannot choose our friends without triggering sanctions, NATO base deployments, or malign-intent ascriptions, then our security is never safe. This inference predicts coalition behavior outside the Western order but does not by itself predict





coalition defections within it. Moving forward, it will be important to see whether the mechanism operates similarly across both subsets of behavior. Returning to the main argument, these five pairs of episodes form a minimum illustrative sample that meets the criteria for selection laid out above. Each pair shares a common feature that pits the universalist articulation of a normatively consequential principle against the asymmetric application of that principle in practice. That structural commonality should, in turn, provoke systematic activation of the five mechanisms hypothesized in the conceptual framework. Other explanations rooted either in variations in material capabilities or in variations in norm diffusion logically cannot predict it.

Objective 4: Deriving falsifiable expectations. The research results found that future empirical work should test whether fairness-asymmetry indicators predict diplomatic misalignment before, or independently of, material-capability indicators. Possible indicators include the timing of official statements, UN voting shifts, sanctions-coalition participation, accession to alternative institutions, local-currency settlement agreements, and public-opinion measures of trust in international organizations.

Counter-cases and boundary conditions. The mechanism does not predict automatic defection. Some states may perceive selective moralism yet remain aligned with Western positions because of security dependence, debt exposure, trade concentration, elite preferences, regime survival calculations, or lack of viable institutional alternatives. Such counter-cases are analytically useful because they distinguish grievance activation from behavioral opportunity.

Discussion

In the interest of fairness, let me distinguish between the article's empirical claims and their interpretation. The claim is that there is a recurrent pattern in the selective-moralism frame, and it seems to matter across diverse cases by tying back to plausible behavioral mechanisms. The interpretation is that selective moralism is contributing to the understanding of trust decay in Western-led institutions. This interpretation could be stronger if it were contrasted systematically with realist and domestic-political explanations instead of being argued as if it were mutually exclusive with them. Structural realism is still a big gorilla in the room. Strategic/external balancing behavior, economic dependence, and domestic political incentives can all constrain leaders' decisions to act on grievance. For that reason, the behavioral story is most convincing to me when it is considered as an intervening mechanism: material shifts create opportunity space, and perceived unfairness in the procedures travels down to influence the perceived legitimacy, timing, and rhetorical cover for defection.

This review should also prompt a re-reading of current geopolitical competition. The behavioral-political-economy perspective argued for here does not contradict the realist emphasis



on material capabilities and structural causes (Mearsheimer, 2019; Schweller, 2018; Allison, 2017), but rather clarifies how structural change becomes translated into ideational and behavioral shifts via its micro-foundations. Three takeaway points follow. First, the trap of selective moralism is a structural condition, not merely a rhetorical gambit. Each additional observation of perceived double standards has lowered the marginal cost for non-Western governments of abandoning or subverting Western institutions and raised the marginal benefit of strengthening ties with alternative powers. The mechanism is self-reinforcing because defections are interpreted — through the lens of selective moralism — as additional evidence that those institutions lack either the will or ability to enforce their own rules universally. Institutional reform efforts that focus on the efficiency or effectiveness dimensions without making credible commitments to even-handedness are therefore destined to further alienate stakeholders (Tyler, 2006; OECD, 2022; Hurd, 2007). Second, trust is a double-edged sword. The foundations of the liberal order have always been predicated on a claim of universal membership, where integrity is therefore the primary source of legitimacy. Once political leaders begin doubting each other's integrity, signals of competence and benevolence are inadequate to rebuild trust because they fail to address the fairness critique head-on. By contrast, non-Western powers' counter-narratives are built around premises of indivisible security, civilizational difference, and multipolarity, all of which assume rejection of limited universality up front and thus benefit from weaker exposure to integrity-based critiques from the outset. The structural conditions that create these asymmetries can be—and have been—reversed historically, but not without substantially changing the universalist claim itself (Reus-Smit, 2007; Wiener, 2018). Third, the dynamic described above matters for the prevention of political violence. Selection effects and brokerage have long been discussed in civil war and intra-state violence literatures as factors enabling those with grievances to mobilize violent means (Krause & Milliken, 2009; Kalyvas, 2019). A similar dynamic operates at the international level. Episodes of perceived selective moralism do not cause wars, but they do change the risk calculus of states by lowering the reputational cost of defecting from institutions meant to mitigate conflict and by conferring relative legitimacy to unilateral or coalitional substitutes. International organizations and agencies tasked with conflict prevention should view fair procedures as a foreign policy asset, not as an ethical concession. The point should come as no surprise to readers familiar with the literature on procedural justice (Tyler, 2006; OECD, 2022; Levi & Stoker, 2000), nor with research on the multiplex nature of order competition in today's world (Acharya, 2018). By focusing on leader-level beliefs and decisions, however, this paper goes one step further in explaining why that competition looks the way it does: legitimacy gaps, transmitted through international institutions and embodied in leaders' beliefs about fairness, cause leaders to make foreign policy decisions that produce outcomes (feedback), reinforcing the





belief. To this extent, my argument here coheres with Adler-Nissen and Zarakol's (2021) account of status politics in world society, but places the causal arrow in a different direction by viewing status less as a fixed quality than as a behavioral response to perceived injustice. My interpretation of the empirical record is also inconsistent with realist expectations of gradual movement along a unidimensional axis of alignment dictated by capability shifts, or liberal expectations of gradual norm convergence as non-Western states increase their interaction density with universal institutions. The account developed here is congruent with the pattern of sudden, clustered defections that persist over time and cut across a variety of regional contexts better than either of those alternatives, though I recognize that the data examined cannot exclude those processes working in parallel with the mechanism proposed. This article has three limitations. The case selection was illustrative, not systematic, and future research should collect structured corpora of official statements, roll-call votes, and survey data to confirm that these claims hold before accepting them as validated. The within-vs.-without contrast obscures interesting gradations in the middle that would be illuminated by distinguishing between regions or coalition blocs within each broad category. Finally, the entire review is based on official statements and documents; it says nothing about how leaders' beliefs about fairness are shaped by or translated to domestic publics, which is a separate but equally important phenomenon. Dedicated experimental, survey-based, and longitudinal research designs could fruitfully pursue each of these questions in ways that complement the conceptual synthesis advanced here.

This qualification changes the strength, not the direction, of the argument. Selective moralism should not be read as a single master cause of order contestation. It is a trust-degrading mechanism that operates alongside power transition, regime interest, economic coercion, domestic coalition politics, and status competition. Its value lies in explaining why states with different ideologies and different material positions can converge around similar accusations of hypocrisy.

Recommendation

1. Specific issue: institutional trust repair. International institutions should not treat double-standard accusations as merely communicative problems. The practical application is to introduce procedural-fairness audits that assess consistency, neutrality, representativeness, and correctability in decision-making. Audits should be public, periodic, and linked to corrective mechanisms; otherwise, they risk becoming another example of stated principle without applied consequence.

2. Specific issue: education and professional training. Curricula in international relations, political science, and global citizenship should include non-Western approaches as analytical



frameworks rather than as peripheral case material. The practical application is to train students, diplomats, and civil servants to distinguish universal legal argument from selective political application. This reduces defensive universalism and improves the capacity to negotiate across contested normative orders.

3. Specific issue: conflict prevention. Early-warning systems should include indicators of perceived procedural unfairness alongside conventional indicators of capability and intent. Practical indicators include recurring hypocrisy frames in official speeches, abrupt changes in UN voting, sanctions-avoidance cooperation, declining trust in international organizations, and coalition participation outside Western-led forums.

4. Suggestions for future research. Future studies should build structured corpora of official statements, roll-call votes, and public-opinion data; compare fairness-asymmetry indicators with material-capability indicators; test whether procedural acknowledgement reduces identity-protective rejection; and examine counter-cases where perceived hypocrisy does not produce diplomatic defection. Experimental, survey-based, and longitudinal designs would allow the mechanism proposed here to be tested rather than only theorized.

References

- Acharya, A. (2018). *Constructing global order: Agency and change in world politics*. Cambridge University Press.
- Adler-Nissen, R., & Zarakol, A. (2021). Struggles for recognition: The liberal international order and the merger of its discontents. *International Organization*, 75(2), 611–634.
<https://doi.org/10.1017/S0020818320000454>
- Allison, G. (2017). *Destined for war: Can America and China escape Thucydides's trap?* Houghton Mifflin Harcourt.
- Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.
- Buzan, B., & Lawson, G. (2015). *The global transformation: History, modernity and the making of international relations*. Cambridge University Press.
- Cooley, A., & Nexon, D. H. (2020). *Exit from hegemony: The unraveling of the American global order*. Oxford University Press.
- Downs, G. W., & Rocke, D. M. (1994). Conflict, agency, and gambling for resurrection: The principal-agent problem goes to war. *American Journal of Political Science*, 38(2), 362–380. <https://doi.org/10.2307/2111408>



- Hafner-Burton, E. M., Haggard, S., Lake, D. A., & Victor, D. G. (2017). The behavioral revolution and international relations. *International Organization*, 71(S1), S1–S31.
<https://doi.org/10.1017/S0020818316000400>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.
- Hopf, T. (2002). *Social construction of international politics: Identities and foreign policies, Moscow, 1955 and 1999*. Cornell University Press.
- Hurd, I. (2007). *After anarchy: Legitimacy and power in the United Nations Security Council*. Princeton University Press.
- Ikenberry, G. J. (2018). The end of liberal international order? *International Affairs*, 94(1), 7–23.
<https://doi.org/10.1093/ia/iix241>
- Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *Cultural Cognition Project Working Paper Series*, Article 164. Yale Law School.
<https://doi.org/10.2139/ssrn.2973067>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kalyvas, S. N. (2019). The landscape of political violence. In E. Chenoweth, R. English, A. Gofas, & S. N. Kalyvas (Eds.), *The Oxford Handbook of terrorism* (pp. 11–33). Oxford University Press.
- Krause, K., & Milliken, J. (2009). Introduction: The challenge of non-state armed groups. *Contemporary Security Policy*, 30(2), 202–220. <https://doi.org/10.1080/13523260903077296>
- Lake, D. A., Martin, L. L., & Risse, T. (2021). Challenges to the liberal order: Reflections on *International Organization*. *International Organization*, 75(2), 225–257.
<https://doi.org/10.1017/S0020818320000636>
- Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. *Annual Review of Political Science*, 3, 475–507. <https://doi.org/10.1146/annurev.polisci.3.1.475>
- Mattern, J. B., & Zarakol, A. (2016). Hierarchies in world politics. *International Organization*, 70(3), 623–654. <https://doi.org/10.1017/S0020818316000126>
- Mearsheimer, J. J. (2019). Bound to fail: The rise and fall of the liberal international order. *International Security*, 43(4), 7–50. https://doi.org/10.1162/isec_a_00342
- OECD. (2022). *Building trust to reinforce democracy: Main findings from the 2021 OECD survey on drivers of trust in public institutions*. OECD Publishing.
<https://doi.org/10.1787/b407f99c-en>
- Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 100(3), 641–672.
<https://doi.org/10.1257/aer.100.3.641>





- Pouliot, V. (2016). *International pecking orders: The politics and practice of multilateral diplomacy*. Cambridge University Press.
- Pytlík-Zillig, L. M., & Kimbrough, C. D. (2016). Consensus on conceptualizations and definitions of trust: Are we there yet? In E. Shockley, T. M. S. Neal, L. M. Pytlík-Zillig, & B. H. Bornstein (Eds.), *Interdisciplinary perspectives on trust: Towards theoretical and methodological integration* (pp. 17–47). Springer.
- Reus-Smit, C. (2007). International crises of legitimacy. *International Politics*, 44(2–3), 157–174.
<https://doi.org/10.1057/palgrave.ip.8800182>
- Schnellenbach, J., & Schubert, C. (2015). Behavioral political economy: A survey. *European Journal of Political Economy*, 40(Part B), 395–417.
<https://doi.org/10.1016/j.ejpoleco.2015.05.002>
- Schweller, R. L. (2018). Three cheers for Trump's foreign policy: What the establishment misses. *Foreign Affairs*, 97(5), 133–143.
- Stuenkel, O. (2020). *The BRICS and the future of global order* (2nd ed.). Lexington Books.
- Tyler, T. R. (2006). *Why do people obey the law?* Princeton University Press.
- Voeten, E. (2021). *Ideology and international institutions*. Princeton University Press.
- Wiener, A. (2018). *Contestation and constitution of norms in global international relations*. Cambridge University Press.
- Zürn, M. (2018). *A theory of global governance: Authority, legitimation, and contestation*. Oxford University Press. <https://doi.org/10.1093/oso/9780198819974.001.0001>

