Research Article

De-Escalatory Institutionalism in the Grey Zone: How Informal Regional Institutions Manage Great-Power Rivalry Where Formal Authority Fails

Junfu Lu[*,a,1]

[a]*UG Y2 student of International Relations, School of Humanities and Social Sciences, Xi'an Jiaotong-Liverpool University*

Compiled on May 16, 2026

Abstract

Why do informal regional institutions sometimes outperform formal universal ones in managing acute security tensions? This article develops the concept of de-escalatory institutionalism to explain how the Association of Southeast Asian Nations (ASEAN) has occupied a distinctive functional niche in South China Sea (SCS) governance — one that formal bodies, particularly the United Nations Security Council (UNSC), are structurally incapable of filling. We argue that institutional effectiveness under conditions of hegemonic contestation cannot be measured by enforcement capacity or legalistic authority alone. Instead, a second dimension of effectiveness — de-escalatory capacity — determines which institutions successfully prevent low-intensity coercive interactions from cascading into open armed conflict. Drawing on process-tracing of two critical junctures (the 2002 Declaration on the Conduct of Parties and the aftermath of the 2016 Permanent Court of Arbitration ruling), we identify three interlocking causal mechanisms: normative enmeshment, ritualized ambiguity, and reputational friction. These mechanisms operate precisely because ASEAN lacks formal enforcement authority, rendering it politically accessible to a revisionist great power that rejects external juridical oversight. The article contributes to theoretical debates on institutional design, informal governance, and escalation management, and develops broader propositions about the conditions under which de-escalatory institutionalism is likely to prevail over formal multilateralism in managing grey-zone competition.

Keywords: *ASEAN; South China Sea; de-escalatory institutionalism; grey-zone conflict; informal institutions; great-power rivalry; norm localization; institutional effectiveness*

1. INTRODUCTION: THE PARADOX OF INFORMAL EFFECTIVENESS

The South China Sea presents international relations theory with an uncomfortable empirical puzzle. Since the early 1990s, the sea has been the site of intensifying territorial contestation, coercive grey-zone operations, large-scale artificial island construction, and repeated near-misses between naval vessels. Yet despite the presence of a revisionist great power — the People's Republic of China — systematically challenging the rules-based maritime order, and despite the complete structural paralysis of the world's most authoritative security institution, the region has not experienced large-scale inter-state war. How should we account for this outcome?

The standard answer offered by liberal institutionalists — that international institutions constrain behavior through binding rules and enforcement mechanisms — does not hold here. The United Nations Security Council (UNSC) has been entirely marginalised from South China Sea governance. China's permanent membership and veto power render it constitutionally incapable of passing substantive resolutions on any dispute in which Beijing is a claimant party. The 2016 arbitral award under the UN Convention on the Law of the Sea (UNCLOS), delivered by a tribunal constituted under Annex VII, was dismissed by China as 'null and void and has no binding force' (Chinese Government Statement, 2016). Formal international law, in this instance, achieved precisely the opposite of its intended purpose: rather than settling the dispute, the legal ruling hardened Chinese intransigence and accelerated coercive gray-zone activity in its aftermath.

And yet conflict has been managed — imperfectly, incompletely, but consequentially. Scholars and policymakers routinely credit the Association of Southeast Asian Nations (ASEAN) with keeping diplomatic channels open, preventing individual crises from escalating, and maintaining what Evelyn Goh (2008) terms a *hierarchical regional order* in which the costs of open armed conflict remain prohibitive. But how, precisely, does ASEAN accomplish this? What mechanisms translate the association's famous institutional 'softness' — its informality, its consensus norm, its non-binding declarations — into actual conflict management? And what are the theoretical implications for our broader understanding of institutional effectiveness and international order?

This article argues that existing theoretical frameworks fail to capture ASEAN's mode of institutional performance because they conceptualise effectiveness exclusively in terms of formal authority, enforcement capacity, and dispute resolution. We propose an alternative concept — de-escalatory institutionalism — to capture a distinct form of institutional effectiveness: the capacity to prevent low-level coercive interactions from crossing thresholds of armed confrontation, without resolving underlying disputes or possessing coercive instruments of its own. De-escalatory institutionalism operates through three interlocking mechanisms — normative enmeshment, ritualized ambiguity, and reputational friction — that collectively raise the diplomatic and reputational costs of escalation for revisionist actors without triggering the confrontational dynamics that formal enforcement tends to generate.

Our argument builds on but departs from several existing theoretical traditions. From the institutionalist literature (Keohane 1984; Koremenos, Lipson, and Snidal 2001), we retain the insight that institutional design matters — but we challenge the assumption that formal legalism represents the summit of institutional development. From the constructivist tradition (Acharya 2004; Finnemore and Sikkink 1998; Wendt 1999), we draw the observation that normative frameworks shape state behavior through socialization — but we specify more precisely the causal mechanisms through which this occurs in a context of acute power asymmetry. From the grey-zone and escalation management literatures (Mazarr 2015; Christensen 2015; Schelling 1960), we adopt the core insight that managing the threshold between

[*]Corresponding author: Junfu.Lu24@student.xjtlu.edu.cn; Tel: 13712702286 (Junfu Lu).

Research article in International Relations, School of Humanities and Social Sciences, Xi'an Jiaotong-Liverpool University.

competition and war is itself a form of security governance — but we connect this to institutional design in ways the strategic studies literature has not fully explored.

The article proceeds as follows. Section II reviews the relevant theoretical literature and identifies the research gap that motivates our intervention. Section III develops the theoretical framework of de-escalatory institutionalism and specifies its three core mechanisms. Section IV outlines our methodology. Section V applies the framework through process-tracing of two critical junctures in SCS governance. Section VI addresses alternative explanations, including U.S. hegemonic deterrence and ASEAN's internal divisions. Section VII draws broader theoretical implications and identifies the conditions under which de-escalatory institutionalism is most likely to prevail. Section VIII concludes.

2. THEORETICAL LANDSCAPE AND RESEARCH GAP

2.1. The Institutionalist Tradition and Its Limits

Mainstream international relations theory has long grappled with the question of when and how international institutions shape state behaviour. The foundational liberal institutionalist position, most systematically developed by Robert Keohane (1984), holds that institutions reduce transaction costs, enhance information, and create iterated interaction environments that make cooperation rational even among self-interested states. Subsequent work on rational design by Koremenos, Lipson, and Snidal (2001) has refined this insight, arguing that states strategically design institutions — their membership rules, scope conditions, flexibility provisions, and enforcement mechanisms — to address specific collective action problems.

These frameworks generate a clear set of expectations about institutional effectiveness: institutions that possess stronger enforcement mechanisms, more clearly defined legal mandates, and broader membership should, ceteris paribus, prove more effective at managing international disputes. By this standard, the UNSC should dramatically outperform ASEAN. It possesses legally binding authority under Chapter VII of the UN Charter, universal membership, and — in principle — enforcement instruments ranging from sanctions to authorised military force. ASEAN, by contrast, operates on the basis of consensus, issues only hortatory declarations, and has no mechanism for compelling member-state compliance, let alone for constraining external powers.

Yet this expectation is systematically violated in the South China Sea context. The institutionalist tradition, as Mearsheimer (1994, p. 7) noted in a different register, struggles to explain why states *comply* with institutions beyond their immediate interest in doing so. In the SCS, China does not comply with UNCLOS as interpreted by the arbitral tribunal, and there is no mechanism capable of compelling it to do so. But it does *participate* — however grudgingly — in ASEAN-centred diplomatic processes. Understanding why requires theoretical tools that the standard institutionalist framework does not currently provide.

2.2. Constructivism, Norm Localization, and the ASEAN Literature

The constructivist tradition offers more purchase on the ASEAN phenomenon. Amitav Acharya's (2004) seminal contribution on *norm localization* — the process by which local actors reconstruct external norms to make them congruent with indigenous frameworks — provides a crucial theoretical resource. Acharya demonstrates that ASEAN's institutional distinctiveness is not mere weakness but a deliberate adaptation of liberal international norms to a regional context in which sovereignty sensitivities are acute and state-building remains incomplete. The 'ASEAN Way' — with its emphases on non-interference, informality, and consensus — is not a deficient copy of Western multilateralism but a locally adapted institutional form with its own logic and limitations (Acharya 2007; Emmers 2014).

Evelyn Goh's (2008) analysis of hierarchical order and strategic *omni-enmeshment* represents the most sophisticated existing account of ASEAN's role in regional order. Goh argues that smaller Southeast Asian states pursue a multi-layered strategy of engaging rising powers diplomatically while relying on U.S. hegemonic deterrence for hard-security assurances. ASEAN, in this reading, is the institutional vehicle through which enmeshment is operationalized — embedding China in webs of diplomatic interaction that raise the costs of coercive behaviour. Goh's framework helpfully distinguishes between hard deterrence and soft engagement, but it does not fully theorise *how* the enmeshment process actually prevents escalation at the level of specific crisis events. It identifies the strategy without specifying the mechanisms.

The hedging literature offers related insights. Kuik Cheng-Chwee (2008) theorises hedging as a strategy that smaller states adopt under conditions of uncertainty about a rising power's long-term intentions, combining elements of engagement and balancing to maintain strategic flexibility. While hedging theory focuses primarily on state strategy rather than institutional design, it underscores the premium that Southeast Asian states place on preserving diplomatic channels — a preference that ASEAN's institutional architecture directly accommodates. Similarly, Bisley (2017) has argued that ASEAN's apparent institutional weakness — its consensus norm and avoidance of legally binding commitments — is in fact a form of organisational *adaptive resilience*, allowing the institution to maintain cohesion across member states with divergent interests and security alignments.

2.3. Grey-Zone Conflict and Escalation Management

A second body of literature, drawn primarily from strategic studies, addresses the phenomenon of grey-zone conflict — coercive activities that fall below the threshold of armed conflict but above the threshold of routine peacetime competition (Mazarr 2015; Brands 2016). In the SCS, China's grey-zone toolkit includes coast guard patrols, maritime militia operations, artificial island construction, and harassment of foreign fishing and survey vessels — all designed to alter the territorial *status quo* while avoiding the kind of direct military engagement that might trigger a U.S. or multilateral response (Fravel 2011; Cronin 2012). Thomas Christensen (2015) has identified the management of grey-zone escalation as one of the central challenges of contemporary U.S. China policy: the risk is not primarily of deliberate war but of inadvertent escalation through crisis mismanagement.

The escalation management literature has largely focused on dyadic state interactions and the role of military signalling, crisis communication, and deterrence thresholds (Schelling 1960; Fearon 1994). It has paid relatively little attention to the role that multilateral institutions might play in managing escalation thresholds — partly because formal institutions are typically considered irrelevant in situations where great-power interests diverge, and partly because the grey-zone literature has developed largely independently of the institutionalist tradition.

This article bridges these two literatures. We argue that informal regional institutions can play a distinctive and undertheorised role in grey-zone conflict management — not by resolving disputes or compelling compliance, but by creating diplomatic environments in which de-escalation is the path of least resistance for all parties. This insight requires a new conceptual framework.

2.4. Informal Institutions and the Research Gap

Recent theoretical work on informal international institutions (Vabulas and Snidal 2013; Raustiala and Victor 2004; Stone 2011) has demonstrated that institutional informality is not simply an absence of formal structure but a distinct institutional type with characteristic advantages and disadvantages. Informal institutions can provide governance in areas where formal institutions are politically blocked, they can serve as 'forum-shopping' venues for states that prefer to

avoid legally binding commitments, and they can generate soft norms that gradually harden into more robust expectations over time. Randall Schweller and Xiaoyu Pu (2011) have further argued that rising powers strategically use informal institutions to challenge existing order while maintaining a veneer of multilateral engagement — a dynamic directly relevant to China's participation in ASEAN-centred processes.

Despite this growing literature, the specific mechanisms through which informal institutions manage grey-zone conflict under conditions of hegemonic contestation remain underspecified. Existing accounts either describe ASEAN's diplomatic role in broadly functional terms or criticise it for failing to resolve underlying disputes — neither of which captures the de-escalatory logic that we identify as central to ASEAN's institutional performance. The research gap is therefore not merely empirical but theoretical: the field lacks a framework that can explain why institutional informality sometimes constitutes an asset rather than a liability in managing acute security tensions.

3. DE-ESCALATORY INSTITUTIONALISM: A THEORETICAL FRAMEWORK

3.1. Reconceptualising Institutional Effectiveness

We propose that existing measures of institutional effectiveness require disaggregation. The literature has predominantly equated effectiveness with two related capacities: dispute resolution (the ability to produce authoritative settlements of contested claims) and compliance generation (the ability to induce state behaviour consistent with institutional rules). Both capacities presuppose a form of authority — either juridical or enforcement-based — that informal institutions characteristically lack.

We argue for a third dimension of institutional effectiveness: ***de-escalatory capacity*** — the ability to prevent ongoing coercive interactions from crossing thresholds of open armed conflict. De-escalatory capacity operates at the level of crisis dynamics rather than dispute settlement; it is concerned not with the ultimate resolution of competing claims but with managing the day-to-day friction that accompanies such competition. This distinction matters because, in a context of acute great-power rivalry, disputes are unlikely to be resolved on any policy-relevant time horizon — the relevant question becomes not 'who wins?' but 'how do we prevent this from becoming a war?'

De-escalatory capacity constitutes a ***functional niche*** that is not filled by formal institutions in the SCS context. The UNSC's authority is premised on great-power consensus; when consensus is absent — as it necessarily is when a permanent member is a direct party to the dispute — the institution is not merely ineffective but actively counterproductive, since failed resolutions signal international disorganization and may embolden revisionist behaviour. Formal legal processes, as the 2016 arbitration demonstrated, can similarly entrench rather than resolve disputes when the losing party possesses both the capacity and the political will to reject the outcome. In such environments, the very features that make informal institutions appear weak — their non-binding character, their ambiguity, their avoidance of zero-sum outcomes — become functional assets.

3.2. Three Causal Mechanisms

Mechanism 1: Normative Enmeshment. Drawing on Goh (2008) and Acharya (2004), we specify normative enmeshment as the process by which an institution embeds a great power in regularized patterns of diplomatic interaction — summits, working groups, multilateral consultations — that generate reputational stakes in continued participation and restrained behaviour. The mechanism operates through what Michael Barnett and Martha Finnemore (2004) call *institutional socialisation*: repeated engagement in norm-governed settings creates expectations, routines, and identity associations that incrementally constrain behaviour. Crucially, enmeshment does not require the great power to accept the institution's normative premises; it requires only that the costs of *exit from* the diplomatic framework are perceived as higher than the costs of *continued participation* in it. ASEAN's dense network of summits, forums, and working groups — including the ASEAN Regional Forum (ARF), the ASEAN Defence Ministers' Meeting Plus (ADMM-Plus), and the East Asia Summit (EAS) — ensures that China's diplomatic calendar is saturated with ASEAN-centred engagement, raising the visibility and reputational cost of any decision to abandon the framework.

Mechanism 2: Ritualized Ambiguity. A second mechanism operates through what we term *ritualized ambiguity* — the institutionalised practice of deferring sovereignty questions while maintaining the form of multilateral engagement. ASEAN's insistence on consensus and its avoidance of binding legal commitments does not represent institutional failure; rather, it functions as a deliberate architecture of productive ambiguity. By refusing to adjudicate sovereignty claims, ASEAN prevents diplomatic encounters from becoming zero-sum confrontations in which China (or any other claimant) must choose between conceding core interests and abandoning the multilateral forum altogether. The resulting dialogues operate, in Schelling's (1960) terms, as *focal points* — shared reference frameworks that coordinate de-escalatory expectations without requiring agreement on substance. The annual ASEAN-China consultations on the Code of Conduct (COC) exemplify this dynamic: they have produced no legally binding outcome in over two decades, yet they have consistently provided a diplomatic space in which both sides can signal restraint, manage bilateral irritants, and prevent specific incidents from triggering broader escalation.

Mechanism 3: Reputational Friction. The third mechanism operates through the *reputational costs* that ASEAN membership imposes on coercive behaviour. Alastair Iain Johnston (2003) has demonstrated that China's participation in multilateral institutions generates genuine socialization effects — not merely strategic mimicry — through processes of persuasion, social influence, and identity management. ASEAN's diplomatic culture creates an environment in which overt coercion generates reputational penalties: accusations of violating the 'spirit' of the DOC, denunciations in ASEAN joint communiqués, and damage to Beijing's narrative of 'peaceful rise.' These penalties do not deter China from grey-zone operations, but they do raise the threshold for actions that would constitute unambiguous departures from professed norms — such as directly confronting ASEAN member-state naval vessels or openly repudiating diplomatic commitments. Reputational friction thus creates *escalation speed bumps* that slow the tempo of crisis escalation even when they cannot prevent coercive activity altogether.

3.3. Boundary Conditions

De-escalatory institutionalism does not operate under all conditions. We identify three boundary conditions that determine when this form of institutional effectiveness is most likely to prevail. First, the revisionist power must retain a preference for continued participation in the diplomatic framework — either because it values the reputational legitimacy that participation confers or because exit would trigger countervailing responses it wishes to avoid. If China were to formally withdraw from ASEAN-centred processes, the de-escalatory mechanism would collapse. Second, the informal institution must maintain sufficient internal cohesion to present a credible collective diplomatic front; severe divisions among member states — as occurred at the 2012 ASEAN Summit, which failed to produce a joint communiqué for the first time in the organisation's history — can signal disorganization that emboldens rather than constrains revisionist behaviour. Third, the grey-zone competition must remain below the threshold at which hard deterrence becomes the dominant security logic — a threshold primarily determined by U.S. strategic commitments. De-escalatory institutionalism operates in the space between diplomatic manage-

ment and military deterrence; if deterrence fails, informal institutions have no fallback capacity.

4. METHODOLOGY: PROCESS-TRACING AND CASE SELECTION

This article employs process-tracing as its primary methodological tool. Process-tracing — the systematic analysis of causal processes through examination of within-case evidence — is particularly well-suited to identifying and validating causal mechanisms, as distinct from establishing correlational patterns across cases (Bennett and Checkel 2015; Mahoney 2015). Our aim is not to demonstrate that ASEAN outperforms the UNSC across a sample of cases, but to trace, at a granular empirical level, how the three mechanisms identified above operated in specific historical episodes.

We focus on two critical junctures in SCS governance, selected because they represent high-stakes moments at which the risk of escalation was acute and institutional responses were consequential: (1) the negotiation and adoption of the 2002 Declaration on the Conduct of Parties in the South China Sea (DOC), which represents the most significant diplomatic achievement of ASEAN-centred SCS management; and (2) the diplomatic aftermath of the July 2016 Permanent Court of Arbitration ruling, which represents the hardest test case for de-escalatory institutionalism, given China's categorical rejection of the legal outcome and the intense pressure it placed on ASEAN cohesion.

For each case, we trace the observable implications of our three proposed mechanisms — asking whether normative enmeshment, ritualized ambiguity, and reputational friction operated in the manner theorised, and whether they generated observable effects on escalation dynamics. Our evidence draws on official diplomatic documents, ASEAN joint communiqués and ministerial statements, Chinese government statements, published accounts by scholar-practitioners with inside knowledge of ASEAN processes (notably Acharya, Emmers, and Goh), and secondary analysis of crisis incidents documented in the open-source literature on maritime security.

We acknowledge three significant methodological limitations. First, this is a single-region, two-case analysis; our causal claims are necessarily provisional and will require validation through comparative cases. Second, the opacity of diplomatic negotiations — particularly within ASEAN's closed-door consultations — means that direct evidence of the mechanisms is often indirect and inferential. Third, we cannot establish a definitive counterfactual: we cannot know with certainty that escalation would have occurred in ASEAN's absence. These limitations are inherent to the study of conflict that does not occur; we address them by triangulating across multiple evidence types and by identifying the empirical implications that would follow from each alternative explanation.

5. EMPIRICAL ANALYSIS: DE-ESCALATORY INSTITUTIONALISM IN PRACTICE

5.1. Critical Juncture I: The 2002 Declaration on the Conduct of Parties

The 2002 Declaration on the Conduct of Parties in the South China Sea represents a paradigmatic example of de-escalatory institutionalism. The declaration emerged from a period of acute tension following China's occupation of Mischief Reef in 1995 — then an unoccupied feature claimed by the Philippines — and subsequent episodes of naval confrontation in the late 1990s. By 2001, ASEAN and China had agreed to negotiate a Code of Conduct that would govern the behaviour of claimant parties; the resulting DOC, signed in November 2002, was deliberately weaker than the binding Code of Conduct that ASEAN had initially sought.

The DOC's weakness — it is non-binding, contains no enforcement mechanism, and merely commits parties to 'exercise self-restraint in the conduct of activities that would complicate or escalate disputes' (Article 5) — has frequently been identified as evidence of ASEAN's institutional failure. We interpret it otherwise. The DOC's adoption demonstrates normative enmeshment at work: China agreed to participate in a multilateral diplomatic framework that formally acknowledged the existence of competing claims and the legitimacy of ASEAN's role as a convening institution. By signing the DOC, Beijing accepted — at least declaratorily — a set of norms including the peaceful resolution of disputes, refraining from inhabiting 'presently uninhabited islands, reefs, shoals, cays, and other features' (Article 5), and conducting cooperative activities in areas such as marine environmental protection and navigational safety. These commitments were not legally binding, but they created reputational stakes: subsequent Chinese actions that violated the spirit of the DOC — including the acceleration of artificial island construction after 2013 — generated diplomatic costs that ASEAN's enmeshment framework helped to *make visible and salient* to a broader regional audience.

Ritualized ambiguity is equally visible in the DOC process. The declaration's deliberate vagueness on the central question of sovereignty — it acknowledges 'the importance of the Declaration on the Conduct of Parties in the South China Sea' without resolving any territorial question — allowed China to participate without conceding its 'nine-dash line' claims, while simultaneously allowing ASEAN claimants (Vietnam, the Philippines, Malaysia, Brunei) to frame their participation as consistent with their own sovereignty positions. This mutual face-saving function, which a binding legal text could not have provided, is precisely what enabled a diplomatically significant outcome to be reached at all. The alternative — insisting on legally binding commitments that China would have rejected — would have produced no agreement and no diplomatic framework, leaving ASEAN-China relations without any normative architecture.

The DOC also illustrates the reputational friction mechanism. In the years following 2002, China repeatedly invoked the declaration as evidence of its commitment to peaceful resolution — deploying ASEAN diplomatic engagement as a constitutive element of its 'peaceful rise' narrative (Zheng 2005). This rhetorical investment created what we term a reputational hostage: the more Beijing publicly committed to the DOC framework, the higher the reputational cost of actions that departed from its normative content. The 2011 cable-cutting incident — in which Chinese maritime surveillance vessels severed the survey cables of Vietnamese vessels in UNCLOS-defined Vietnamese Exclusive Economic Zone — triggered precisely this dynamic: Vietnam was able to invoke the DOC framework in condemning Chinese behaviour, and China's response, through senior officials including Dai Bingguo, emphasised continued commitment to the declaration's principles, representing a diplomatic de-escalation signal even as grey-zone operations continued.

5.2. Critical Juncture II: The 2016 PCA Ruling and Its Aftermath

The aftermath of the July 2016 Permanent Court of Arbitration ruling constitutes the hardest empirical test for our framework. The tribunal's award — ruling that China's nine-dash line historical claims had no legal basis under UNCLOS and that China had violated the Philippines' exclusive economic zone rights — was the most significant juridical development in SCS governance since the adoption of UNCLOS itself. China's response was categorical: the government declared the ruling 'null and void' and explicitly rejected the legitimacy of the arbitral process. International observers anticipated a severe escalation of Chinese coercive activity in the weeks following the ruling.

The escalation did not materialise in the form anticipated. While China continued grey-zone operations, it did not formally occupy additional Philippine-claimed features or confront Philippine naval vessels in ways that would have constituted a clear departure from its prior behaviour. We argue that ASEAN's de-escalatory mechanisms played a significant role in this outcome — though one whose causal impor-

tance is easily underestimated because it operated through absence of escalation rather than through visible diplomatic activity.

The critical dynamic was the intense pressure placed on ASEAN as an institution in the days following the ruling. China lobbied intensively to prevent ASEAN from issuing a joint statement that acknowledged the ruling's legal significance. The resulting ASEAN foreign ministers' statement of July 2016 — which mentioned the ruling only tangentially and did not call on China to comply — was widely criticised as evidence of Chinese influence over ASEAN and the organisation's failure to uphold international law. From the perspective of dispute resolution effectiveness, this criticism is entirely warranted.

From the perspective of de-escalatory institutionalism, however, the outcome requires a different interpretation. The preservation of ASEAN cohesion — even on terms highly favourable to China — maintained the diplomatic framework through which de-escalatory signalling could continue. A fractured ASEAN that publicly condemned China's legal non-compliance would have destroyed the very institutional architecture on which de-escalatory institutionalism depends, leaving no multilateral venue for crisis management. The pragmatic compromise reflected the organisation's internal logic: maintaining the institutional framework itself, even at the cost of a maximalist legal position, was judged by the majority of ASEAN members to be preferable to institutional breakdown.

Crucially, the ASEAN framework enabled the subsequent Duterte government's decision to pursue bilateral diplomatic engagement with China — a decision that produced a temporary de-escalation of Philippine-Chinese tensions and the informal reopening of Philippine access to Scarborough Shoal. Whatever the long-term strategic consequences of this diplomatic accommodation, it demonstrates that ASEAN's de-escalatory architecture created the diplomatic space within which bilateral de-escalation became possible. A context in which ASEAN had publicly fractured over the ruling, or in which the Philippines had been diplomatically isolated as a result of the arbitration, would have made such outcomes far less likely.

The 2016 episode also illustrates the boundary conditions identified in our theoretical framework. The near-failure of ASEAN cohesion — the reported division between Cambodia and Laos, on one side, and Vietnam and the Philippines, on the other — demonstrates how internal fragmentation can undermine de-escalatory institutionalism. When the association cannot maintain even a diplomatically ambiguous collective position, the normative enmeshment and reputational friction mechanisms lose their efficacy: China faces no collective diplomatic front and can manage dissenting members bilaterally, exploiting its vastly superior economic leverage.

6. ALTERNATIVE EXPLANATIONS AND SCOPE CONDITIONS

6.1. The Hegemonic Deterrence Thesis

The most powerful alternative explanation for SCS stability attributes it not to ASEAN but to U.S. hegemonic deterrence. In this reading, the absence of large-scale armed conflict reflects China's rational calculation that an overt military challenge to SCS claimants would risk triggering U.S. military intervention — a risk China is not yet prepared to accept given the current military balance. This argument is most forcefully articulated by Goh (2008) in her analysis of hierarchical order, and it has significant empirical support: U.S. carrier strike group deployments, freedom of navigation operations (FONOPs), and security commitments to the Philippines under the U.S.-Philippines Mutual Defense Treaty all constitute credible deterrent signals.

We do not dispute that U.S. deterrence constitutes a necessary condition for SCS stability at the level of large-scale armed conflict. Our argument is rather that deterrence alone is insufficient to explain the absence of escalation at the level of grey-zone incidents — precisely the level at which most SCS friction occurs. U.S. deterrence credibly constrains China from seizing a Philippine island by force or blockading Vietnamese ports; it does not constrain China from deploying coast guard vessels to harass Vietnamese fishing boats or from dredging sand around contested features. Grey-zone operations are specifically designed to exploit the gap between deterrence thresholds and day-to-day competitive behaviour — a gap that hard deterrence is by design poorly suited to fill.

De-escalatory institutionalism operates precisely in this gap. ASEAN cannot deter Chinese grey-zone operations, but it can create diplomatic incentives and reputational costs that shape the tempo and manner of those operations. The two explanations are therefore not competitive but complementary: U.S. deterrence sets the upper bound of Chinese coercive behaviour (preventing large-scale armed attack), while ASEAN's de-escalatory mechanisms manage the lower register of grey-zone friction. Analytically separating these two functions is essential for correctly attributing causal weight to each.

6.2. The ASEAN Failure Thesis

A second alternative explanation emphasises ASEAN's failures rather than its successes. Critics note that China has dramatically expanded its SCS presence since 2002 — constructing artificial islands, installing military infrastructure, and deploying maritime militias — all while continuing to participate in ASEAN-centred diplomatic processes. In this reading, ASEAN's diplomatic framework serves China's interests by legitimising continued engagement while providing cover for ongoing coercive activity; it is not a constraint on Chinese behaviour but a diplomatic instrument of it.

This critique has significant merit and must be engaged seriously. There is no question that ASEAN has not prevented China's grey-zone expansion. Our framework explicitly does not claim that de-escalatory institutionalism resolves disputes or stops all coercive behaviour; we claim only that it manages escalation thresholds. The appropriate empirical test is not whether ASEAN has stopped Chinese grey-zone activity, but whether ASEAN's diplomatic framework has demonstrably affected the escalation dynamics of SCS crises — whether specific incidents have been de-escalated through mechanisms consistent with our framework.

The evidence reviewed in Section V supports a qualified affirmative answer. The DOC's reputational dynamics constrained the form if not the fact of Chinese grey-zone operations; ASEAN's post-2016 diplomatic management preserved channels through which bilateral de-escalation became possible. These are modest claims, but they are theoretically significant: they demonstrate that de-escalatory capacity constitutes a real and consequential form of institutional effectiveness, even in the absence of dispute resolution or enforcement capacity.

7. THEORETICAL IMPLICATIONS AND GENERALIZATION

7.1. Implications for Institutional Theory

Our analysis carries several implications for broader debates in institutional theory. Most fundamentally, it argues for the disaggregation of institutional effectiveness into at least three analytically distinct dimensions: dispute resolution capacity, compliance generation, and de-escalatory capacity. Current theory, by conflating these dimensions, systematically undervalues the governance contribution of informal regional institutions and overvalues the role of formal legal authority in security contexts characterised by great-power rivalry.

The concept of de-escalatory institutionalism also contributes to the growing literature on institutional design under uncertainty (Koremenos 2016; Jupille, Mattli, and Snidal 2013). Our analysis suggests that *deliberate institutional vagueness* — far from representing design failure — can constitute a strategic asset in contexts where the parties to a dispute have fundamentally incompatible views of the underlying legal and historical facts. Ritualized ambiguity is not the same as institutional paralysis; it is a specific design feature that enables continued engagement without requiring substantive agreement.

The boundary conditions we identify also contribute to the literature on institutional resilience and fragility (Rixen and Viola 2015). De-escalatory institutionalism is most effective when the revisionist power retains a preference for institutional participation — a condition that holds as long as the reputational benefits of engagement outweigh the costs of the normative constraints associated with it. When this condition is violated — when a revisionist power concludes that exit from the diplomatic framework is costless or advantageous — informal institutions lose their de-escalatory function and may even become counterproductive by providing false assurances about the management of underlying tensions.

7.2. Generalizability: Beyond the South China Sea

The theoretical framework developed here is not limited to the ASEAN-SCS context. De-escalatory institutionalism is likely to be relevant wherever three conditions obtain simultaneously: (1) a formal universal institution is structurally blocked from acting by great-power veto or jurisdictional constraints; (2) grey-zone competition operates below the threshold of deterrence; and (3) an informal regional institution maintains sufficient cohesion and great-power participation to function as a diplomatic framework. These conditions describe not only the SCS but several other contemporary governance challenges — including management of Russian grey-zone activity in Eastern Europe (where the OSCE has played an analogous de-escalatory role), competition in the Arctic (where the Arctic Council's informal architecture manages great-power friction despite the absence of enforcement capacity), and emerging cyber conflict governance (where informal multilateral frameworks may eventually develop de-escalatory functions comparable to those identified here).

The South China Sea case also generates a broader theoretical proposition about the relationship between institutional formality and effectiveness under hegemonic contestation. We suggest that the conventional wisdom — that stronger institutional authority produces better governance outcomes — reverses under conditions in which the relevant great power is a claimant party to the dispute, possesses veto authority or jurisdictional opt-out options in formal institutions, and retains the capacity to exit informal institutions without severe sanctions. In these conditions, formal institutional authority is an asset for compliance among small and medium powers but becomes a liability for engaging revisionist great powers. Informality, paradoxically, enables the very engagement that makes de-escalation possible.

7.3. The Limits of De-Escalatory Institutionalism

We conclude this section with a candid assessment of the limits of our framework. De-escalatory institutionalism is not a solution to the underlying disputes that drive grey-zone competition. It manages symptoms rather than causes. In the SCS context, the fundamental tension between China's expansive territorial claims and the legal rights of other claimants under UNCLOS remains entirely unresolved — and may be moving in China's favour, as the incremental consolidation of artificial island positions changes the material facts on the water. ASEAN's de-escalatory function may, in the long run, provide cover for a slow-motion revision of the territorial status quo that no party is willing to contest through force.

This suggests a temporal horizon problem inherent in de-escalatory institutionalism: mechanisms that successfully manage short-term escalation thresholds may, by preserving the diplomatic framework, inadvertently legitimate a long-term revision of order that is ultimately more destabilising than the short-term escalation they prevented. The appropriate institutional response to this dilemma — whether to insist on formal legal outcomes at the risk of institutional breakdown, or to preserve informal engagement at the cost of normative ambiguity — is a question that de-escalatory institutionalism, by design, defers rather than resolves.

8. CONCLUSION

This article has developed the concept of de-escalatory institutionalism to explain an enduring puzzle in South China Sea governance: why an informal, non-binding, and frequently divided regional institution has managed acute security tensions more effectively than the world's most formally authoritative security body. We have argued that effectiveness in grey-zone conflict management requires a distinct analytical dimension — de-escalatory capacity — that cannot be captured by existing measures focused on dispute resolution or compliance generation. ASEAN's de-escalatory capacity operates through three interlocking mechanisms: normative enmeshment, ritualized ambiguity, and reputational friction. These mechanisms are causally enabled, not weakened, by ASEAN's institutional informality.

The broader theoretical contribution is a revision of how institutional effectiveness should be assessed under conditions of hegemonic contestation. Where formal institutions are structurally blocked by great-power rivalry, and where grey-zone competition operates below the threshold of deterrence, the relevant question is not which institution possesses the most authority but which institution can sustain the diplomatic engagement that makes de-escalation possible. Institutional informality — traditionally treated as a deficiency — can become a functional asset in precisely these conditions.

Three implications deserve emphasis for future research and policy. First, scholars of international institutions should develop more refined measures of de-escalatory effectiveness, enabling systematic comparison across regions where informal institutions manage great-power friction. Second, comparative case studies of the OSCE, the Arctic Council, and emerging cyber governance frameworks should test whether the mechanisms identified here operate across different strategic contexts. Third, policymakers designing institutions for grey-zone conflict management should consider whether the deliberate incorporation of ritualized ambiguity — rather than the maximalist pursuit of legally binding outcomes — might better serve de-escalatory functions in contexts of acute great-power rivalry.

The South China Sea will not be resolved by ASEAN. But in a world where formal institutions are increasingly gridlocked by great-power competition, understanding how informal institutions maintain the diplomatic infrastructure of de-escalation may be among the most consequential contributions that international relations scholarship can make to the governance of a turbulent international order.

REFERENCES

Acharya, A. (2004). How ideas spread: Whose norms matter? Norm localization and institutional change in Asian regionalism. *International Organization*, 58(2), 239–275.

Acharya, A. (2007). The emerging regional architecture of world politics. *World Politics*, 59(4), 629–652.

Barnett, M., & Finnemore, M. (2004). *Rules for the world: International organizations in global politics*. Cornell University Press.

Bennett, A., & Checkel, J. T. (Eds.). (2015). *Process tracing: From metaphor to analytic tool*. Cambridge University Press.

Bisley, N. (2017). ASEAN's mid-life crisis. *Asian Affairs*, 48(3), 381–401.

Brands, H. (2016). Paradoxes of the gray zone. *Foreign Policy Research Institute*, February.

Buszynski, L. (2012). The South China Sea: Oil, maritime claims, and U.S.–China strategic rivalry. *The Washington Quarterly*, 35(2), 139–156.

Christensen, T. J. (2015). *The China challenge: Shaping the choices of a rising power*. W. W. Norton.

Cronin, P. M. (2012). *Cooperation from strength: The United States, China and the South China Sea*. Center for a New American Security.

Emmers, R. (2014). ASEAN's search for neutrality in the South China Sea. *Asian Journal of Peacebuilding*, 2(1), 61–77.

Fearon, J. (1994). Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model. *Journal of Conflict Resolution*, 38(2), 236–269.

Finnemore, M., & Sikkink, K. (1998). International norm dynamics and political change. *International Organization*, 52(4), 887–917.

Fravel, M. T. (2011). China's strategy in the South China Sea. *Contemporary Southeast Asia*, 33(3), 292–319.

Goh, E. (2008). Great powers and hierarchical order in Southeast Asia: Analyzing regional security strategies. *International Security*, 32(3), 113–157.

Gould, M., & Rablen, M. D. (2017). Reform of the United Nations Security Council: Equity and efficiency. *Public Choice*, 173(1–2), 145–168.
Ikenberry, G. J. (2011). *Liberal leviathan: The origins, crisis, and transformation of the American world order*. Princeton University Press.
Johnston, A. I. (2003). Socialization in international institutions: The ASEAN way and international relations theory. In G. J. Ikenberry & M. Mastanduno (Eds.), *International relations theory and the Asia-Pacific* (pp. 107–162). Columbia University Press.
Jupille, J., Mattli, W., & Snidal, D. (2013). *Institutional choice and global commerce*. Cambridge University Press.
Keohane, R. O. (1984). *After hegemony: Cooperation and discord in the world political economy*. Princeton University Press.
Koremenos, B. (2016). *The continent of international law: Explaining agreement design*. Cambridge University Press.
Koremenos, B., Lipson, C., & Snidal, D. (2001). The rational design of international institutions. *International Organization*, 55(4), 761–799.
Kuik, C.-C. (2008). The essence of hedging: Malaysia and Singapore's response to a rising China. *Contemporary Southeast Asia*, 30(2), 159–185.
Mahoney, J. (2015). Process tracing and historical explanation. *Security Studies*, 24(2), 200–218.
Mazarr, M. J. (2015). Mastering the gray zone: Understanding a changing era of conflict. *United States Army War College Press*.
Mearsheimer, J. J. (1994). The false promise of international institutions. *International Security*, 19(3), 5–49.
Permanent Court of Arbitration. (2016). *In the matter of the South China Sea arbitration (PCA Case No. 2013-19): Award*. PCA.
Raustiala, K., & Victor, D. G. (2004). The regime complex for plant genetic resources. *International Organization*, 58(2), 277–309.
Rixen, T., & Viola, L. A. (Eds.). (2015). *Institutional change in international political economy*. MIT Press.
Schelling, T. (1960). *The strategy of conflict*. Harvard University Press.
Schweller, R., & Pu, X. (2011). After unipolarity: China's visions of international order in an era of U.S. decline. *International Security*, 36(1), 41–72.
Stone, R. W. (2011). *Controlling institutions: International organizations and the global economy*. Cambridge University Press.
Vabulas, F., & Snidal, D. (2013). Organization without delegation: Informal intergovernmental organizations (IIGOs) and the spectrum of intergovernmental arrangements. *Review of International Organizations*, 8(2), 193–220.
Wendt, A. (1999). *Social theory of international politics*. Cambridge University Press.
Zha, W. (2024). Ethnic politics, political security, and the selective application of the non-interference principle within ASEAN. *Asian Journal of Peacebuilding*, 12(2), 231–255.
Zheng, B. (2005). China's 'peaceful rise' to great-power status. *Foreign Affairs*, 84(5), 18–24.

A. AUTHOR'S NOTE ON THEORETICAL CONTRIBUTION, ORIGINAL ESSAY WEAKNESSES, SIMULATED REVIEWER REPORT, AND FUTURE EXTENSIONS

A.1. Main Theoretical Contribution

The article's principal theoretical contribution is the concept of de-escalatory institutionalism — a framework that reconceptualises institutional effectiveness as a multi-dimensional phenomenon including dispute resolution capacity, compliance generation, and de-escalatory capacity. It challenges the assumption that formal authority is a prerequisite for institutional effectiveness in security governance, demonstrating that informality can be a functional asset under conditions of hegemonic contestation. The three-mechanism framework (normative enmeshment, ritualized ambiguity, reputational friction) provides causal precision that existing accounts of ASEAN's role — which tend to be descriptive or broadly functional — have lacked. The article also contributes a set of generalisable boundary conditions and broader propositions about when de-escalatory institutionalism prevails over formal multilateralism.

A.2. Weaknesses in the Original Essay

Framing as comparative effectiveness: The original essay asked 'Why has ASEAN been more effective than the UNSC?' — a framing that implied a like-for-like comparison between institutions with fundamentally different mandates, memberships, and logics. This produced a category error that undermined analytical precision from the outset.

Absence of causal mechanisms: The essay identified several relevant empirical phenomena (grey-zone management, norm localization, enmeshment) but did not specify how these translated into causal outcomes. Claims about ASEAN's effectiveness remained descriptive rather than mechanistic.

Theoretically underdeveloped: The essay drew on a relatively narrow set of theoretical resources (Mearsheimer, Acharya, Goh, Emmers) without engaging with the broader institutionalist, informal institutions, escalation management, or grey-zone literatures. The theoretical contribution was implicit rather than explicitly developed.

Inadequate engagement with alternative explanations: The hegemonic deterrence explanation was acknowledged but not systematically engaged; its relationship to ASEAN's diplomatic role was left conceptually unclear.

Thin empirical analysis: The process-tracing was referenced as a methodology but not executed in detail; specific causal sequences were asserted rather than traced through documentary evidence.

Limited generalizability: The essay remained confined to the SCS context without developing propositions that might travel to other cases of informal institutional governance under great-power rivalry.

A.3. Simulated International Organization Reviewer Report

Manuscript evaluation — simulated in the style of International Organization:

This manuscript addresses an important and undertheorised puzzle in international relations: the governance performance of informal regional institutions in contexts of great-power competition. The central concept of de-escalatory institutionalism is original, theoretically precise, and clearly motivated by a genuine gap in the literature. The disaggregation of institutional effectiveness into three dimensions — dispute resolution, compliance generation, and de-escalatory capacity — represents a valuable analytical contribution that should prompt productive debate.

The three-mechanism framework is the manuscript's strongest element. Normative enmeshment, ritualized ambiguity, and reputational friction are clearly defined, theoretically grounded in existing scholarship, and empirically illustrated with appropriate granularity. The discussion of ritualized ambiguity is particularly innovative and deserves further development in future work.

Major revisions recommended: (1) The process-tracing in Section V would benefit from more systematic engagement with primary diplomatic documents — particularly ASEAN joint communiqués and Chinese government statements — to strengthen the evidentiary basis for mechanism identification. (2) The boundary conditions in Section III.C are well-identified but could be developed into more precise scope conditions with clearer empirical indicators. (3) The relationship between de-escalatory institutionalism and existing hedging theory (Kuik 2008; Goh 2008) should be more explicitly theorised: is de-escalatory institutionalism a form of institutional hedging, or a distinct phenomenon?

Minor revisions: (1) The section on generalizability (VII.B) gestures at comparative cases (OSCE, Arctic Council, cyber governance) without sufficient development; these comparisons should either be developed or deferred to future work. (2) Some claims about the 2016 PCA aftermath would benefit from additional citation of primary documentation.

Overall assessment: The manuscript makes a significant theoretical contribution and should be considered for publication subject to major revisions addressing the points above. With appropriate revision, it has the potential to make an important contribution to debates on institutional effectiveness, informal governance, and grey-zone conflict management.

A.4. Suggestions for Future Extensions and Comparative Case Studies

OSCE and the Ukraine Crisis (2014–present): The Organisation for Security and Cooperation in Europe played a de-escalatory role in the early stages of the Donbas conflict through its Special Monitoring Mission. Examining whether normative enmeshment, ritualized ambiguity, and reputational friction operated in this context — and whether they ultimately failed when Russia concluded that exit was costless — would provide an important comparative test of the boundary conditions identified here.

The Arctic Council under Great-Power Competition: The Arctic Council's hybrid architecture — with observer states, indigenous peoples' representatives, and great-power members — provides another test case for de-escalatory institutionalism in a context of growing Sino-Russian-American competition over Arctic resources and shipping lanes.

ASEAN in the Taiwan Strait: Extending the analysis to examine whether ASEAN's de-escalatory mechanisms have any purchase on cross-strait tensions — and whether their absence in this context confirms the boundary conditions of the framework — would be theoretically productive.

Temporal Dynamics of De-Escalatory Institutionalism: A longitudinal study examining whether ASEAN's de-escalatory capacity has changed over time — as Chinese material power has increased relative to the U.S. and as domestic political pressures within ASEAN member states have grown — would test the framework's durability and identify conditions under which de-escalatory institutionalism erodes.