

High, Dry and Allied: The Ethics of Breaching the Collective Defence Duty

David Rubin

Independent Researcher, corresponding address: david.rubin1002@outlook.com

Abstract

This article adopts a normative approach to one of the most consequential issues of alliance management: the question of if and when an ally is justified in breaching its collective defence duty (CDD). An ally's CDD constitutes its formal obligation to militarily defend its ally in the event that the latter experiences an armed attack. Drawing on both normative political theory and contemporary International Relations scholarship on alliance credibility and security dilemmas, the article first considers and rejects two opposing but equally extreme views on this question. Next, the article proposes what it calls the Moderate View as a more persuasive response. The Moderate View draws on Glenn Snyder's concept of the alliance security dilemma to argue that Ally X is justified in breaching its CDD if it is reasonable for Ally X to believe that there is a high likelihood of its imminent entrapment by Ally Y, and if Ally X cannot reduce this likelihood by signalling its concerns to Ally Y. Two distinct arguments are then provided in support of the Moderate View. In the final section, the Moderate View is defended against three objections. The article demonstrates how empirical concepts from IR can inform normative theorising while providing practical ethical guidance for alliance policymakers.

Keywords: alliances, ethics, collective defence, security dilemma, war

First published online on 28 January 2026

Introduction

More than three years into Russia's full-scale invasion of Ukraine, and following Donald Trump's return to the US presidency in early 2025, European NATO members find themselves confronting a question that has haunted allied states throughout history: Will their most powerful ally honour its commitment to defend them in their hour of need? While Article 5 of the NATO treaty legally obligates all members to assist any ally suffering an armed attack, Trump's repeated expressions of scepticism towards the alliance – including past suggestions that the United States might not defend allies failing to meet defence spending targets – have intensified European concerns about potential American abandonment (Jonah 2025). This anxiety reflects a perennial dilemma of alliance politics, which take place against the backdrop of an anarchic international system in which interstate agreements remain unenforceable, and the intentions of even close allies can never be completely known (Mearsheimer 2001; Waltz 1979). In contemporary alliance politics, it is not hard to find other cases where the risk is present in the mind of a state vis-à-vis its allies: for example, Australia, Japan and the Philippines vis-à-vis mutual ally the United States against the backdrop of a rapidly rising and ever more assertive China. If these threatened states failed to receive the military assistance that their crucial US ally has a treaty duty to provide to them, they would find themselves in an exceedingly dangerous situation.

Yet despite these cases and the continuing salience of the issue in the international security domain, and despite the large amount of *positive* – i.e. descriptive, explanatory or predictive – work on the subject in international relations (IR), IR scholars have shown no interest in exploring the *normative* implications of an ally breaching its obligation to defend its ally from attack (though some rare exceptions are Kunertova 2017 and Rubin 2023). The study of alliances has evolved considerably over recent decades. What might be termed 'first generation' alliance research focused primarily on capability aggregation models, wherein states ally in order to pool military resources against common threats (Walt 1987; Waltz 1979), and on security-autonomy trade-offs, wherein asymmetric allies exchange military protection for policy influence (Morrow 1991).

This body of research, while foundational, has increasingly ceded ground to 'second generation' research that often employs rational choice and bargaining models in order to examine alliance formation and maintenance through the theoretical lenses of strategic interaction, credible commitment problems and incentives for domestic political survival. Todd Sandler's (2006) influential work on collective action, in particular, has demonstrated that alliance dynamics extend well beyond simple capability aggregation to encompass complex problems of burden-sharing, free-riding and collective goods provision. Similarly, a robust literature on alliance credibility has examined how domestic political institutions

– particularly democratic electoral mechanisms – affect the perceived trustworthiness and reliability of security commitments (Fearon 1994; Gaubatz 1996; Schultz 1999; Weeks 2008). A key insight of this literature is that any given allied state's decision to enter a conflict in defence of its treaty ally will of necessity be endogenous to the alliance commitment itself. Put otherwise, the dynamics that drive the process of alliance formation (i.e. who commits to ally with whom) will tend to shape subsequent conflict behaviour (i.e. who takes action to defend whom) in ways that first generation theories did not fully capture.^[^7] Constructivist scholarship, in turn, has at once critiqued and enriched rationalist insights and findings of this sort by highlighting how alliances can constitute security communities built upon shared identities, norms and practices rather than purely upon each state's interest-based and security-driven calculations of the net benefits of collective defence (Adler and Barnett 1998; Risse 1996; Wendt 1999). Constructivists have above all emphasised that states may honour alliance commitments due to what James March and Johan Olsen (1998) famously term a 'logic of appropriateness' – i.e. due to a preference for acting in accordance with internalised norms – rather than to what they call a 'logic of consequences'.

What unites these diverse scholarly approaches – i.e. first generation and second generation, rationalist and constructivist – is their uniform focus on positive inquiry rather than normative inquiry. Similarly, scholars working in the international political theory (IPT) subfield of political philosophy, employing a normative approach to issues in international politics, have shown little to no interest in alliances *tout court*, let alone in the ethical dimensions of an ally's abrogating its collective defence duty.

This article aims to address these overlapping gaps in the IR and IPT literatures by adopting an explicitly normative approach to alliance management, and specifically to the question of when, if ever, a state is justified in breaching its duty to defend its ally. The present article does not attempt to supplant or compete with rationalist or constructivist approaches to alliances. Rather, it addresses a gap that exists across all of them: the systematic normative analysis of alliance obligations. While rationalist approaches excel at explaining when and why states form alliances and predicting when commitments will be honoured or breached, they typically bracket questions of moral justification. When is a state morally justified in breaching its collective defence duty? What ethical principles should govern this decision? These are not questions that rational choice models are designed to answer, nor are they questions that have been addressed by the extensive literatures on alliance credibility and political survival. Similarly, while constructivists demonstrate that states often perceive themselves as bound by genuine obligations to allies, they do not typically engage in the prescriptive project of determining what those obligations should entail or when they may permissibly be violated. This is where normative political theory makes its dis-

tinctive contribution. The IPT subfield takes normative questions – questions of right and wrong, justice and injustice, permissibility and impermissibility – as its main focus. IPT scholars have extensively analysed the ethics of war (McMahan 2009; Walzer 2015), humanitarian intervention (Pattison 2012) and distributive justice (Rawls 1999), yet alliances have received virtually no sustained normative attention despite their centrality to international security politics. This article aims to fill that gap by bringing the tools of normative analysis to bear on collective defence obligations.

Such an approach is worth undertaking for several reasons. To begin with, a state's formal obligation to militarily aid its allies carries with it a host of interesting and important ethical questions, of which I will focus on one in particular. Furthermore, a normative approach to collective defence duties can both complement and enrich the enormous positive literature on alliance formation and alliance management by helping to facilitate a greater scholarly awareness of the potential moral complexities and dilemmas embedded in alliance politics. The approach taken here can be characterised as critical in at least two senses: It critically examines taken-for-granted assumptions about alliance obligations, refusing to accept either that such obligations are absolute or that they are merely epiphenomenal to power politics; and it draws on critical concepts from the IR literature – in particular, Glenn Snyder's alliance security dilemma (1984) – to inform normative theorising, thereby bridging empirical insights about alliance dynamics with ethical reasoning about obligations. Finally, the analysis of alliances from a normative perspective can furnish political actors and policymakers with critical information about the ethics of a given strategy or policy that is adopted, or a given action that is taken, by their state in relation to the defence of an ally. In short, a normative approach to alliances, and particularly to the question of if and when an ally is justified in abrogating the collective defence duty, would be a valuable addition to scholarship and policy research on the subject. Extrapolating from the foregoing discussion, the specific contributions of this article can be understood as being three in number. It provides, first of all, the first systematic normative account of, and framework for, evaluating when the breach of an alliance's collective defence duty (CDD) may be justified. In addition, the article demonstrates how empirical concepts from positive research on alliances within IR – most notably, the alliance security dilemma – can productively inform the aforementioned normative theorising, anchoring the latter in the often unpredictable and perilous complexities of real-world alliance politics and international security.

The article proceeds as follows. In the first section, I outline the methodological approach employed in this article. In the second section, I define an alliance and discuss the nature of the collective defence duty that arises from alliance membership. In the third and fourth sections, I present and then reject two extreme

views – what I call the Idealist View and the Hobbesian View – on the question of whether the collective defence duty can ever justifiably be breached by an ally. In the fifth section, I introduce what I call the Moderate View on this question and then go on to support the view with two distinct arguments. In the sixth section, I defend the Moderate View against three objections. A brief final section concludes.

Methodological approach

Before examining competing views on when CDD breach may be justified, it is necessary to clarify the methodological approach employed in this article. This clarification is especially important given that modern alliance scholarship spans multiple methodological traditions, including large-n quantitative studies, game-theoretic formal modelling and qualitative comparative case studies.¹ These methodologies, however, grounded as they are in empirical hypothesis testing and aiming as they do at positive inquiry, are clearly inappropriate for undertaking the kind of normative inquiry pursued in this article. The main aim of that inquiry is not to explain why allied states do or do not honour their CDD, nor to predict under what conditions such breaches will occur. Rather, the primary aim is to develop a normative framework for determining when allied states are morally justified in breaching the CDD. The most appropriate methodology for undertaking this aim is philosophical argumentation. This methodology involves clearly defining concepts, employing those concepts in the formulation of principles, developing arguments in support of these principles, considering counterarguments to the latter and objections to the principals themselves, refining the principles in light of these critiques, and testing the normative and practical tenability of the principles against concrete empirical cases, which may be either actual or hypothetical in nature.² It is this general methodological approach, which has become standard within normative political theory and applied ethics, that structures and guides this article's use of hypothetical alliance scenarios grounded in real-world alliance contexts. These hypothetical cases serve several purposes that are worth elaborating.

The use of hypothetical cases allows, first of all, for greater clarity by testing normative principles under relatively controlled conditions in which relevant

1 For large-n quantitative studies, see for example Leeds *et al.* (2002). For formal modelling, see for example Smith (1995). For qualitative case studies, see for example Pressman (2008).

2 This approach is sometimes broadly and loosely described as the method of 'reflective equilibrium', which was first formulated by John Rawls (1971) and which instructs us to begin with our considered moral judgments about particular cases, then to formulate general principles to explain these, and then finally to move back and forth between judgments and principles – revising either as needed – until we are able to attain a coherent, mutually supporting system in which our specific beliefs and abstract principles are in stable alignment.

variables can be isolated and manipulated. For example, by constructing scenarios in which the entrapment of an ally is clearly imminent and the possibility of signalling is either unavailable or has demonstrably failed, we are arguably in a better position to assess whether these conditions truly support a possible moral intuition that CDD breach is warranted. Having said this, it is worth stressing that the hypothetical scenarios employed in this article are not arbitrary or overly hygienic thought experiments whose details and dynamics are wholly disconnected from reality.³ Rather, they are constructed with the aim in mind of reflecting real-world alliance relationships (e.g. France-Greece and China-North Korea), existing geopolitical tensions (e.g. Eastern Mediterranean disputes, Korean Peninsula security), plausible conflict dynamics that account for regional power configurations, and pertinent decision-making constraints facing allied states. In this respect, the scenarios function similarly to hypothetical cases in applied ethics that, while not describing actual events, capture the essential features of the ethical dilemmas under investigation. A second purpose served by the use of hypothetical scenarios is that these scenarios facilitate the exploration of boundary conditions and hard cases that may not have clear historical precedents. While there certainly are historically documented cases of alliance abandonment (e.g. France's failure to assist Russia in 1904–05 and, debatably, the Rio Pact's failure to assist Argentina in 1982), it remains true that detailed evidence about the reasoning and circumstances surrounding these breaches is often unavailable or contested. Hypotheticals allow us to specify precisely the conditions under which breach occurs, making the normative analysis more precise. A third purpose served, finally, is that by opting to construct and normatively examine fictive scenarios rather than cases that have actually occurred, it becomes easier to avoid triggering the political allegiances and biases of both author and reader and establish greater objectivity and transparency. Relatedly, it also becomes easier to avoid potential diplomatic sensitivities that might arise from analysing contemporary allies' actual contingency plans or suggesting that particular allies might breach their commitments. Given that the article's analysis could in theory help inform contemporary policy debates, it seems prudent to avoid being perceived as making insinuations concerning the future intentions of named states.

Having described the article's normative methodological approach, the limitations of such an approach should be acknowledged. First, and as implied in the introduction's discussion of positive versus normative inquiry, normative theory-building by definition neither can nor aspires to generate empirically testable predictions about state behaviour. The Moderate View articulates the conditions under which CDD breach is morally justified. It does not, however,

³ See the sharp critique of such hypotheticals, and the framework for evaluating hypotheticals generally, that is offered in Thaler (2018).

predict whether any one member of a given alliance will actually breach its CDD under those conditions or whether that member will be motivated by the Moderate View's normative logic when doing so. Both of these (empirical) questions, in order to be undertaken, require a (positive) methodology that is very different from the (normative) one used in this article. Second, the hypothetical cases employed in the article must necessarily simplify complex geopolitical realities. Real-world alliance decisions inevitably involve a large array of overlapping factors and considerations, including incomplete information, bureaucratic politics and domestic political pressures, none of which can be fully captured in stylised scenarios. Those scenarios, to reiterate a point discussed just above, are chiefly intended to isolate key morally relevant empirical variables for normative analysis. They are not in any way designed to (nor could they possibly) capture the full complexity of actual intra-alliance decision-making.

Alliance preliminaries

I define an alliance as *a bilateral or multilateral institution formed between states and grounded in a formal agreement mandating, under one or more specified circumstances, the deployment or nondeployment of military force by Ally X vis-à-vis Ally Y (with Ally Y representing a vector of all current allies of Ally X)*.⁴ This definition includes the alliance types of offense pact, defence pact, consultation pact, neutrality pact and nonaggression pact.⁵ Importantly, however, the argument developed in future sections applies mainly to defence pacts. Defence pacts commit Ally X to help militarily defend Ally Y in the event of an unprovoked armed attack against Ally Y by a specified or unspecified third state. These pacts tend to be the alliance type that, among scholars, analysts and the general population, is most closely associated with the term 'alliance', possibly due to the enduring prominence of the multilateral defence pact NATO. Hence, when I use the term 'alliance' in what follows, I mainly have in mind defence pacts. Although with appropriate modifications the argument could be made applicable to the other alliance types, in its present form it often proves inapplicable to them. This is because the content of the collective defence duty (see just below) largely assumes a defence pact, and also because Snyder's alliance security dilemma, which is so central to the argument advanced in the sixth section, is less operative in the context of these other types.

With this definition of an alliance spelled out, it is necessary to conceptualise what I have been calling the 'collective defence duty' (hereafter 'CDD'). The CDD, as the concept is used throughout this article, denotes *the formal obligation of Ally X to provide military assistance to Ally Y in the event that Ally Y is the target of an armed attack*. In Hohfeldian terms, the CDD functions essentially as a claim

⁴ For alternative definitions, see Leeds (2020: 6); Snyder (1997: 4); Walt (1987: 12).

⁵ See Leeds *et al.* (2002) for detailed descriptions of these types.

right possessed by Ally Y vis-à-vis Ally X in regard to the military defence of Ally Y in the contingency of an attack on the latter, or (equivalently) as the absence of a liberty right that would otherwise (i.e. if Ally X and Ally Y were not allied) be possessed by Ally X in regard to the same action and contingency (Wenar 2022). As such, the CDD can be said to constitute the deontic foundation of an alliance agreement and the element of that agreement that provides it with its substantive distinctiveness as a treaty under international law. The NATO treaty, for example, contains its CDD in its famous Article 5:

The Parties agree that an armed attack against one or more of them in Europe or North America shall be considered an attack against them all, and consequently they agree that, if such an armed attack occurs, each of them, in exercise of the right of individual or collective self-defense recognized by Article 51 of the Charter of the United Nations, will assist the Party or Parties so attacked by taking forthwith, individually, and in concert with the other Parties, such action as it deems necessary, including the use of armed force, to restore and maintain the security of the North Atlantic area. (North Atlantic Treaty 1949)

International law, of course, has its own rules concerning the circumstances under which a legal obligation may be breached.⁶ Yet the CDD, besides being a legal obligation, also carries with it an interesting and important ethical component, entailing as it does: a solemn promise to undertake a future action; the prospective use of large-scale deadly force in other-defence; the objective entrustment by one (state) agent of its military security to the military protection of another (state) agent; and the subjective trust placed by one (state) agent in the future willingness of another (state) agent to militarily defend it from attack. It would seem to follow that if Ally X were to breach the CDD and so fail to defend Ally Y in Ally Y's hour of need, then Ally X would be morally blameworthy in some way – that is, unless Ally X had a valid justification for failing to fulfil its obligation. The remainder of this article is occupied with the question of whether such a justification does indeed exist and, if it does, under what circumstances it might be available to Ally X as a valid means of avoiding blameworthiness for a breach of the CDD. I start, in the next two sections, by considering and rejecting two extreme views on these interconnected questions. I then introduce and defend a third, more pragmatic view in the sixth section.

6 The most authoritative view on the conditions allowing for the legally permissible breach of treaty terms is contained in Part V of the Vienna Convention on the Law of Treaties. See 'Vienna Convention' (1969).

The idealist view on breaching the CDD

Let us begin with what I call the Idealist View on breaching the CDD. This straightforward view can be summarised as follows: *Ally X is never justified in breaching the CDD in regard to the provision of military assistance to Ally Y in the event that Ally Y has been attacked.* As can be seen, the Idealist View categorically denies that there can ever be a justification for breaching the CDD, so that Ally X will *always* be blameworthy for committing such a breach, no matter the circumstances surrounding the breach or the deleterious consequences for Ally X of upholding its commitment. This uncompromising view is grounded in a very strict and static conception of international ethics and interstate rights and duties, according to which an act or omission is morally wrong across all international contexts and in spite of any *prima facie* mitigating factors. Hence, under the Idealist View, it is not possible for any such contingent context or factor to weaken the deontic force of the CDD, including contexts and factors such as the existence of an overwhelming strategic interest on the part of Ally X in withholding military assistance to Ally Y, the high risk of significant casualties among Ally X's intervening combat personnel or the expectation of intense diplomatic blowback from the international community against Ally X. To take a recent empirical example, the failure of five members of the Collective Security Treaty Organization (CSTO) – i.e. Kazakhstan, Kyrgyzstan, Tajikistan, Belarus and (especially) Russia – to militarily assist sixth member Armenia in its 2020 war with Azerbaijan renders these five members morally blameworthy under the Idealist View, even if their strategic interests undoubtedly favoured nonsupport and even if support would likely have generated for them some number of human casualties and have greatly damaged their diplomatic relations with Azerbaijan (Deyermond 2018; Krivosheev 2021). Put simply, then, the Idealist View is rigid and unconditional in its reading of the CDD; as a matter of principle, it permits no exceptions when it comes to upholding and fulfilling the CDD.

Although the Idealist View has the virtue of prescriptive clarity and unambiguity, these qualities are bought at the heavy price of prescriptive nuance and flexibility once the alliance is confronted with the inherent complexity and danger of real-world collective defence scenarios. Put another way, the Idealist View is unpersuasive as a basis for interpreting the content and scope of the CDD because its interpretation of the CDD would prove, by its very starkness and strictness, deeply unreasonable in some real-life collective defence contexts. All such contexts, of course, unfold against the precarious backdrop of systemic anarchy, though some are characterised by more extreme near-term insecurity than others. At a minimum, a state that finds itself in such a radically insecurity-ridden context must, before pursuing any ancillary goal like collective defence, conduct its foreign policy so as to preserve its own fundamental military security and,

thus, maintain at a satisfactory level its probability of future survival as a state (Waltz 1979). A state that failed to conduct its foreign policy in this way would seem profoundly misguided in its preference ranking of foreign policy priorities. It would seem so misguided not only from a strategic but also arguably from an ethical and legal standpoint, since a state is made up of citizens whose basic welfare and even lives would be put at serious risk by a foreign policy that placed other priorities above those of state survival and fundamental military security, and since the right of (state) self-defence would appear to necessarily override the duty of collective defence as a matter of both international ethical norms and customary international law. One such misplaced priority would, at least under certain circumstances, be the defence of an ally that has been attacked or threatened with attack. If the state that has attacked Ally Y poses a legitimately catastrophic threat to the survival or military security of Ally X, then it is far from evident that a breach of the CDD by Ally X remains categorically unjustified. This is especially true if the threat to Ally X by the aggressor state would be carried out only in the event that Ally X were to militarily intervene in the conflict.

Let us take, for example, the case of the United States–South Korea mutual defence pact (Roehrig and Heo 2018). Suppose first that minor power South Korea has chosen to accept an enormous and potentially existential risk in order to help defend its major power ally in a conflict between the latter and China. China has signalled that, if South Korea does intervene in the conflict, it will execute an immediate and massive nuclear reprisal against numerous counterforce and countervalue targets on the South Korean mainland. South Korean and US intelligence analysts have determined this threat to be highly credible. At the same time, China has conspicuously abstained from making an analogous threat against the US mainland or overseas territories. Accordingly, by choosing – as the Idealist View would demand – to quite literally bite the bullet and fully satisfy its obligations under the CDD, South Korea would be not only drastically, but also (given its basically negligible contribution to the defence of a military superpower and the disproportionate risk surrounding its participation compared to that surrounding US participation) pointlessly and perversely, lowering its prospects of near-term survival. In doing so, South Korea would not appear to be acting in either a strategically rational or – with regard to the welfare of its 52 million citizens,⁷ most of whose lives would be at stake as a consequence of their state's intervention – a morally responsible fashion. In at least this case, then, South Korea – or any similarly situated state – does seem *pro tanto* justified in committing a one-time breach of the CDD. Moreover, it is also likely to seem justified in other, comparably extreme collective defence cases, and perhaps even in some cases that are somewhat less extreme. But if this is true, then the categorical and

7 Assuming, of course, that these citizens have not explicitly consented to be killed.

unconditional Idealist View cannot be the most cogent interpretation of the CDD. Instead, the view's rigidity and dogmatic lack of consideration for the many extreme contingencies that can and do occur within the current international system make it appear strikingly implausible, and so untenable, as a normative framework for conceptualising the deontic relations that exist between allies under anarchy.

The Hobbesian View on breaching the CDD

We now consider the Hobbesian View on breaching the CDD, which is the inverse of the Idealist View. According to the Hobbesian View, whose straightforwardness is equal (if opposite) to that of its rival: *Ally X is always justified in breaching the CDD in regard to the provision of military assistance to Ally Y in the event that Ally Y has been attacked*. As can be seen, the only difference between the content of the Hobbesian View and that of the Idealist View is the presence of 'always' (versus 'never') in the former. The Hobbesian View thus asserts that, while the CDD may technically exist as a normative corollary of the formal act of alliance, it can be justifiably violated at the discretion of Ally X, for any reason whatsoever and without Ally X incurring any blameworthiness thereby. The Hobbesian View can endorse this radical understanding of allied obligations because it is grounded in a conception of international politics according to which rights, duties and the possibility of ethics itself are, particularly in the security realm, strictly subordinate to egoistic considerations of state power and interest, along with the conflictual dynamics generated by the interstate interaction of these considerations. Crucially, the power-and-interest considerations themselves are alleged to be a function of the innate unpredictability and peril of the anarchic system, which socialises states to behave in a manner that maximises power and actualises interests at the expense of all other considerations, thereby optimising chances of survival. To return to the CSTO example, the Hobbesian View would evaluate the failure of the other five CSTO members to militarily aid sixth member Armenia as completely justified, provided that these allies were acting so as to maximise their power and actualise their interests in order to optimise their chances of survival.⁸ Simultaneously, it would also have evaluated these five allies as justified if they had, in fact, chosen to satisfy their CDD obligations towards Armenia, provided that these allies were acting with precisely the same motivations. What the Hobbesian View categorically denies, therefore, is that the CDD can ever be more than a normative epiphenomenon of the rational pursuit of power and interest, as the latter comprises the only feasible and operative 'norm' of state action in international politics and, by extension, alliance politics.

8 The empirical assumptions of the Hobbesian View thus partly overlap with those of certain realist conceptions of international politics; even so, I do claim that all or even most realists would subscribe to the Hobbesian View itself as a *normative* conception of alliance obligations. See Mearsheimer (2001); Morgenthau [1948] (2006); Waltz (1979).

If the Idealist View is hobbled by its radical rejection of a state's ultimate right to prioritise its own survival over the survival of an ally and to protect and preserve its fundamental military security even at the expense of an ally's security, the Hobbesian View exhibits the opposite defect: namely, the elevation of state survival, accomplished through the egoistic pursuit of power and interest, to a sort of categorical imperative in light of which the CDD (along with any other norm of interstate relations) loses all normative force. Such an imperative remains implausible even if one accepts the Hobbesian View's implicit premise that each state (including alliance members) has vis-à-vis other states (including its allies) *only* the duty to optimise its survival chances by maximising its power and actualising its interests. It remains implausible because the premise itself does not entail that an ally is never unjustified in breaching the CDD, since the acceptance of a robust obligation to defend an ally *in at least some collective defence contexts* may, *under at least some circumstances*, be the optimal means of maximising an ally's power and actualising its interests. Indeed, what is probably the leading theoretical approach to explaining alliance formation, the 'capability aggregation' model, assumes that states allying against either objective power or subjective threat believe exactly this.⁹ Similarly, the competing 'security-autonomy trade-off' model retains power and interest considerations in its explanation of alliance formation but argues that these considerations often work their effects through an inter-ally exchange of military security for policy autonomy (Morrow 1991).

Nonetheless, there is good reason to think that the aforementioned premise is false and that states do have duties besides those of survival-oriented power maximisation and interest actualisation. Certainly, this is by far the dominant view within the normative IPT literature, among contemporary philosophers specialising in international ethics and global justice. Indeed, even interstate war itself is seen by the vast majority of IPT scholars, not to mention the vast majority of state leaders and policymakers, as rigorously constrained by sets of reciprocal rights and duties that bind belligerent states both before and during armed conflict, as the rich scholarly tradition of just war theory amply demonstrates.¹⁰ What is more, positive IR scholars adopting a constructivist approach have shown that states can, and not infrequently do, perceive themselves to be bound together with their formal allies in relationships of mutual identification, trust and obligation, enabled and reinforced by processes of ally-level socialisation and norm internalisation (Adler and Barnett 1998; Barnett 1996; Hemmer and Katzenstein 2002; Johnston 2001; Risse 1996; Wendt 1999). The outcome can sometimes, if not

9 The capability aggregation model predicts that states will ally in order to additively pool their capabilities and thereby achieve synergistic efficiency gains in defence provision. See Walt (1987); Waltz (1979).

10 For a detailed overview of modern just war theory, see Lazar (2020).

always, be a cohesive and self-sustaining security community that, because it is built upon shared values, rules and practices, makes possible an (at least partially) collectivised conception of external threats, along with an equally collectivised conception of the duty to deter and defend against such threats. Alexander Wendt's influential 'cultures of anarchy' framework (1999) distinguishes between Hobbesian, Lockean and Kantian cultures of anarchy, corresponding roughly to international orders characterised by enmity, rivalry and friendship, respectively. The Hobbesian View implicitly assumes that all interstate relations, including those between allies, conform to a Hobbesian culture of anarchy where egoistic self-help is the only operative logic. Yet empirical evidence suggests that many contemporary alliances – particularly NATO, the US bilateral defence pacts with Japan, South Korea and Australia, and even the European Union (EU)¹¹ – exhibit characteristics of Lockean or even Kantian cultures wherein restraint, reciprocity and genuine obligation play important roles (Adler 2008; Risse 1996). Examples of existing alliances displaying this degree of deontic 'thickness' include, most notably, the heavily institutionalised NATO and the aforementioned US bilateral defence pacts. Furthermore, there is growing evidence that regime type affects not only alliance credibility but also normative conceptions of alliance obligations. In democracies, the public is able to exercise accountability through electoral mechanisms over leaders who fail to honour their international commitments (Fearon 1994; Gaubatz 1996; Schultz 1999). What it is critical to stress here in the context of a normative analysis is that such accountability can arguably give rise not only to a strategic incentive to honour the aforementioned commitment, but also to a moral obligation to do this. The obligation in question is plausibly grounded in democratic legitimacy and popular sovereignty. Conversely, in the case of authoritarian regimes whose leaders face much weaker accountability mechanisms (or none at all), the normative status of alliance commitments may be different. It may not, however, necessarily be weaker, as clientelist relationships and personalistic ties may generate alternative bases for moral obligation among elites (Svolik 2012; Weeks 2008). Finally, there is even some recent experimental evidence indicating that majorities of the public *within* allied states are conscious of possessing weighty moral obligations to defend their states' allies and would support their state doing so if these allies were attacked (Tomz and Weeks 2021).

Taken together, what all of this normative and empirical work suggests is that the Hobbesian View is very likely mistaken in its almost nihilistic conception of intra-alliance ethics and, in particular, its stringently realpolitik stance on breaching the CDD. Alliances are not merely temporary marriages of convenience that parties may dissolve at will whenever strategic calculations shift. At least some

¹¹ Since 2009, pursuant to Article 42.7 of the Treaty of Lisbon, the EU has been (among its other organisational identities) a multilateral defence pact.

alliances embody genuine normative commitments grounded in shared values, mutual identification, public authorisation and internalised obligations. The challenge, then, is to develop a normative framework that takes these obligations seriously while also recognising the legitimate constraints that anarchy and self-help imperatives place on states' ability to honour commitments under all circumstances. This is precisely what the Moderate View, developed in the next section, aims to accomplish.

The Moderate View on breaching the CDD

Having rejected both the Idealist View and the Hobbesian View as unpersuasive, in this section I propose a compromise position that is located somewhere between these two extremes and that, I argue, yields a more plausible answer to the question of if and when an ally is justified in contravening the CDD. I call this compromise position the Moderate View. In order for the Moderate View to be properly understood, however, I must first introduce Snyder's influential model of the alliance security dilemma from whose logic and concepts the view draws. In several works, Snyder (1984, 1997) builds on the classic concept of the security dilemma in IR by devising a parsimonious model of the risks and dangers embedded in security relations between allies.¹² He calls this model the alliance security dilemma. According to Snyder, this dilemma is endemic to the relationship between allies in an anarchic system and is driven by the two interlocking risks of *entrapment* and *abandonment*. In essence, entrapment occurs when a state is dragged by its ally into an objectively reckless and *de facto* offensive (if speciously defensive) war that, while strongly desired by the ally, is patently contrary to the entrapped state's interests. Entrapment happens as a direct result of the ally's being both emboldened by an expectation of military support from the entrapped state and inclined to exploit the latter's collective defence commitment.¹³ Abandonment, meanwhile, occurs when a state is deserted by its ally in

¹² The classic security dilemma occurs when purely defensive measures taken by State A to enhance its own security (for example, an arms buildup) render State B less secure because State B cannot be sure of the purely defensive (as opposed to offensive) motivations behind these measures. State B thus takes reciprocal defensive measures to increase its own security, causing State A to in turn reciprocate, causing State B to again reciprocate, *ad infinitum*. Hence an insecurity spiral is unintentionally initiated that leaves both states not only less secure than at the outset, but also poorer on account of the futile rounds of military spending. See Jervis (1978).

¹³ Importantly, this version of entrapment is quite different from how the term is used in the domestic legal context. Precise definitions of domestic legal entrapment abound among legal scholars and philosophers, but that of Stitt and James, which posits four conditions for entrapment to have occurred, is typical. It includes four characteristics: first, a law-enforcement agent *plans* a particular crime; second, the agent *induces* the target to commit it; third, the agent *arrests* the target for having committed it; fourth, *counterfactual condition*: if it were not for the agent's actions, then the . . . crime

either peacetime or wartime: i.e. either before the state has been threatened or attacked by the alliance's adversary or after this has taken place. Abandonment therefore includes, while also being conceptually broader than, the narrower act of breaching the CDD, which can only take place once a threat has been issued or an armed attack implemented against the state.

Critically, there is an inverse relationship between entrapment and abandonment: As the risk of one decreases, the risk of the other *ipso facto* increases, and vice versa. This inverse relationship produces the dilemmic aspect of the alliance security dilemma. Hence, when a state, in order to avoid being entrapped by its ally in an undesired military confrontation or conflict, deliberately distances itself from that ally's military aims, policies or actions and so creates (in the ally's mind) extreme doubt as to its own reliability, that state thereby pushes its ally towards abandonment of itself and, potentially, realignment with an ostensibly more dependable and supportive security partner. Conversely, when a state, in order to avoid being abandoned by its ally if and when the former confronts a military threat or attack, deliberately and unconditionally supports its ally's military aims, policies and actions and so creates (in the ally's mind) extreme confidence in its own reliability, that state thereby pushes its ally towards entrapment of itself in an undesired military confrontation or disastrous military conflict. Entrapment and abandonment, then, are obverse sides of the same coin.

With this brief summary of the alliance security dilemma provided, the Moderate View can be stated as follows: *Ally X is sometimes justified in breaching the CDD in regard to the provision of military assistance to Ally Y in the event that Ally Y has been attacked – namely, Ally X is justified in doing so if (1) it is reasonable for Ally X to believe that there is a high likelihood of its imminent entrapment by Ally Y and (2) Ally X cannot reduce this likelihood by clearly signalling to Ally Y its concerns.* Something to notice here before examining the jointly sufficient conditions in (1) and (2) is that the Moderate View's 'sometimes justified' verdict on CDD infringement positions it between the 'never justified' and 'always justified' verdicts of the Idealist View and Hobbesian View, respectively. It should be emphasised that 'sometimes' here implicitly includes, apart from the conditions in (1) and (2), circumstances in which Ally X's very survival is threatened or in which its fundamental military security is at stake. Put differently, the Moderate View assumes that Ally X's survival and fundamental military security are not necessarily at risk in the collective defence context at hand. This means that the conjunction of (1) and (2) amounts to a separate and additional justification,

would not have been committed by the target (1984, 114). As will readily be seen, this domestic legal version of entrapment is formally and substantively distinct from the alliance security dilemma version. I will therefore leave it aside while noting only that a comparative analysis of the two versions on both a descriptive and normative level could make for an interesting topic for future research.

independent from that which might be generated by the aforementioned existential risks,¹⁴ for breaching the CDD.

Now let us look at the two conditions, (1) and (2). (1) stipulates that there must be a reasonable belief on the part of Ally X that a high likelihood of imminent entrapment exists. The italicised words require some clarification. ‘Reasonable’ means that any similarly situated state in Ally X’s position could be counted on to reach, in good faith and based on objective evidence, roughly the same conclusion. ‘Belief’ means that Ally Y need not actually be planning to entrap Ally X in a conflict with a third state, but only that Ally X perceives Ally Y’s policy or behaviour towards either itself or the third state as strongly suggesting this outcome. ‘Imminent’ means that the perceived risk of entrapment must fall squarely in the near term and not the medium or long term – i.e. it must be concrete and impending, not remote and hypothetical. (2), meanwhile, stipulates that there must also be an inability on the part of Ally X to reduce the imminent entrapment risk that it reasonably believes exists by means of clearly signalling its worries to Ally Y. ‘Inability’ means that Ally X must, in good faith, have made its best effort to signal its worries to Ally Y while ultimately failing, despite this effort, to alter the policy or behaviour of Ally Y that is seen as entailing entrapment. ‘Clearly’ and ‘signalling’ together mean that Ally X must have transmitted its concerns to Ally Y, via the appropriate diplomatic channels and at the appropriate diplomatic/ministerial level, in a fashion that allows for both the candid, unambiguous and respectful communication of these concerns and the opportunity for Ally Y to issue a timely response thereto. Finally, as already alluded to, these two conditions are only jointly sufficient for justification: If (1) is satisfied but (2) is not, then Ally X is not justified in breaching the CDD. So, if Ally X reasonably believes that there is a high likelihood of imminent entrapment of itself by Ally Y, yet has made either no effort at all or an inadequate effort to signal its concerns to its ally, then according to the Moderate View, Ally X is not justified in breaching its CDD obligations in the event that Ally Y is attacked.

We now move to the normative substance of the Moderate View. Why, one might ask, should we think that a state is justified in abstaining from militarily defending its ally (i.e. justified in breaching the CDD) due purely to the fact that conditions (1) and (2) obtain in a given alliance scenario? In what follows, I offer two distinct arguments in support of this core claim of the Moderate View. The first of these arguments holds that Ally X is intrinsically justified in breaching the CDD due to its reasonable belief in Ally Y’s imminent entrapment of it. My main claim here is that a high likelihood of entrapment, reified in the form of a reasonable belief and coupled with a good faith effort at signalling concerns, frees Ally X from its duty to defend Ally Y because entrapment itself is a serious breach of Ally Y’s collective defence obligation and of Ally X’s right not to have its own collective defence obligation be

¹⁴ Recall that the Idealist View denies that such a justification can be generated.

exploited in a way that unduly endangers itself. The entrapment of Ally X by Ally Y contravenes Ally Y's alliance obligations *vis-à-vis* Ally X because it entails, as mentioned above, the exploitation by Ally Y of Ally X – or, more exactly, the exploitation of Ally X's formal promise to defend Ally Y in the event that the latter is attacked. Such exploitation contradicts the very spirit of a military alliance, which is based on the transparent and reciprocal exchange of collective defence commitments as well as on some nontrivial baseline of mutual trust and respect between members. As a consequence, by knowingly manipulating Ally X's promise of prospective assistance with the aim of compelling Ally X to participate in an unnecessary conflict that will recklessly imperil Ally X's and its own security, Ally Y is wronging Ally X and undermining the normative foundation of their formal agreement. Indeed, in the absence of this agreement, Ally Y would not be emboldened to embark on such an irresponsible campaign to begin with.

To frame the above points slightly differently, Ally Y is using the alliance as a means for which it was neither explicitly nor implicitly designed: i.e. as an instrument for permitting it to pursue an aggressive and negligent foreign policy, blatantly in contempt of its ally's welfare and wishes. By doing so, Ally Y can also be said to be using Ally X itself as merely a means towards the achievement of a purely self-interested and strategically valuable objective, and even if the process itself of achieving this objective is expected to grossly compromise Ally X's security and harm its fundamental interests. In this way, Ally Y is infringing the foundational deontological principle, first formulated by Kant [1785] (1993), that it is wrong to treat an agent as a mere means towards some given end. Kant, of course, had human persons in mind, but if states can on some ontological level be considered international persons – or even just as emergent agglomerations of individual human agencies – then the principle may well be, *mutatis mutandis*, meaningfully applicable.¹⁵ Whether or not a state might be considered akin to a person, though, it remains true that Ally Y is using its Ally X in a manner to which the latter has never consented, or is ever likely to consent, to being used.

On account of Ally Y's responsibility for this markedly unjust treatment of Ally X, it is plausible that Ally Y thereby forfeits its right to receive the military assistance that Ally X would otherwise, pursuant to the CDD, owe to it – that is, Ally Y is liable to suffer the harm of not being defended.¹⁶ This forfeiture though should best be seen as only temporary: Ally X's CDD obligation *vis-à-vis* Ally Y would be restored were Ally Y to begin respecting the legitimate concerns signalled by Ally X and from thereon work to cease (or drastically curtail) its entrapping conduct.

¹⁵ On the state as an international person, see Wendt (2004).

¹⁶ This formulation has been influenced by Jeff McMahan's 'responsibility account' of liability to be killed in war. McMahan (2009, Ch. 5/2011) views such liability as being grounded in a combatant's moral responsibility for posing an unjust threat.

Nevertheless, until this reversal of conduct happened, Ally X would be justified in breaching the CDD as a reciprocal and proportionate response to Ally Y's wrongdoing. The provisional outcome for the alliance as a whole, therefore, would be the complete suspension (if not cancellation) of Ally X's collective defence obligation towards its ally. In sum, then, the argument moves from the first, empirical premise of Ally X's reasonable belief in its imminent entrapment by Ally Y to the conclusion of Ally Y's temporary loss of its right to be defended and (equivalently) Ally X's justified violation of the CDD. The argument makes this move by virtue of a second, normative premise: the exploitative and agreement-breaching wrongness of Ally Y's entrapment of Ally X.

In its current form, the foregoing argument remains somewhat abstract. To better appreciate the argument, it will be helpful to consider an empirically informed hypothetical illustration. Let us take the case of the recently formed (in late 2021) mutual defence pact between France and Greece. This alliance is widely considered to have been created to deter and defend against shared French and Greek rival Turkey, and specifically against the increasingly assertive and acquisitive Turkish naval presence in the Eastern Mediterranean (Psaropoulos 2021; Sokou 2021). Let us imagine that Greece has entrapped France in a naval conflict with Turkey due to a known desire on the part of a new ultranationalist government in Athens to, first, reestablish Cypriot sovereignty over Northern Cyprus (occupied and controlled by Turkey since 1974) and, second, militarily pursue enosis (political union) of Greece with Cyprus – a long-sought ambition of many ethnically Greek ultranationalists in both states. Greece plans to stealthily provoke a Turkish naval attack on a Greek-flagged drillship that is surveying for natural gas deposits in disputed Aegean Sea waters. France receives credible intelligence on Greek intentions. It perceives the prospective conflict as not only deeply contrary to its own interests and brazenly contradictory of the alliance's spirit and formal aims, but also tremendously detrimental to European and Middle Eastern security and stability. As a result, France is seriously considering breaching its CDD if Greece were to court such an unnecessary, self-interested and destructive conflict with Turkey. France communicates its entrapment concerns to Greece through high-level diplomatic channels and, when this proves fruitless, a French-requested phone call between heads of state in which the French president warns the Greek prime minister against perverting or exploiting the alliance relationship as a tool for guaranteeing its 'defence' in a de facto offensive territorial war. Nonetheless, these communications and warnings from its French ally notwithstanding, Greece's ultranationalist regime signals its determination to press on with its belligerent foreign policy.¹⁷

¹⁷ This scenario is constructed for analytical purposes independent of overlapping NATO obligations that both France and Greece hold. While both are NATO members – which would create additional considerations in reality – this hypothetical examines

Given Greece's plan to entrap France, the imminence of this entrapment, and the imperviousness of the Greek leadership to France's legitimate protests, the Moderate View would evaluate France's intended breach of the CDD as justified in this case, unless and until Greece reverses its entrapping policy or behaviour. According to the first argument, this breach is permissible because Greece has wilfully perverted and exploited its mutual defence relationship with France and betrayed the latter's trust for narrowly egoistic gain. Greece has thereby disregarded French interests and placed French security in grave danger for no necessary or compelling reason. It follows that Greece has forfeited its right to be militarily defended by France in the former's impending irredentist war with Turkey.

Now let us consider the second argument for the Moderate View. This argument holds that Ally X is *instrumentally* justified in breaching the CDD due to its reasonable belief in its imminent entrapment by Ally Y. In this case, my main claim is that a desire on the part of Ally Y for Ally X to uphold the CDD, combined with a reasonable belief on the part of Ally Y that there is a high likelihood of imminent breach by Ally X were Ally Y to entrap it, together make the prospect of Ally Y's entrapment of Ally X less likely. In other words, assuming legitimate entrapment concerns and a good faith but futile effort by Ally X to signal these concerns to Ally Y, the former is justified in refusing to militarily assist Ally Y because the consequences of this conditional refusal, if understood and anticipated by Ally Y, would have positive, anti-entrapment consequences both for Ally X and Ally Y in the near term and for the alliance generally in the long term. The positive consequences for Ally X would be not only its near-term non-entrapment, but also its ability to continue reaping the security-enhancing benefits of collective defence: benefits that Ally Y would also continue to reap, given that it could continue to count on Ally X upholding its CDD obligations. The positive consequences for the alliance as a whole would be a more credible deterrence capacity and (were deterrence to fail) a more formidable defence capacity vis-à-vis shared adversaries, with both consequences flowing from enhanced intra-alliance harmony and cohesion. All of this assumes, again, that Ally Y understands and anticipates the probable consequences of attempting to entrap Ally X, information that could be relatively easily communicated by Ally X to Ally Y *ex ante*, whether during the alliance formation process or after the alliance has been created. In short, then, impending breach of the CDD by Ally X would act as an effective means of dissuading Ally Y from undertaking pro-entrapment policies or actions in the first place, or of pressuring it to stop these once initiated, resulting in positive outcomes for allies and alliance both.¹⁸

the bilateral treaty obligations in isolation to test the normative principles at stake.

18 It should be evident that, in this second argument, it is (instrumentally valuable)

Like the first argument, the second argument could, for the sake of greater clarity, benefit from a hypothetical illustration that is grounded in a real-world alliance scenario. Let us take the case of the China–North Korea mutual defence pact (Albert 2021; Vu 2021). This asymmetric alliance is principally used by China as a vehicle for influencing North Korea’s foreign policy, and by North Korea as a vehicle for protecting itself from external attack (via Chinese power) and as an economic lifeline (via Chinese trade and aid). Before examining this scenario in detail, it is important to acknowledge how regime type and power asymmetry affect alliance dynamics in ways relevant to entrapment risks and CDD obligations. The China–North Korea alliance is characterised by two features that distinguish it from the symmetric, democratic alliances often emphasised in alliance literature. The first feature is that both allied states are authoritarian regimes whose leaders do not face electoral accountability for alliance decisions. The second feature is that the alliance is highly asymmetric, with China possessing vastly greater military and economic power than does North Korea. These two features of the alliance have implications for how we should understand entrapment risks and CDD breach justifications. In terms of the shared authoritarian regime type of the allies, it is known from the work of scholars like Jessica Weeks (2008, 2012) that autocratic leaders face a different set of credibility constraints than do democratic leaders. As discussed briefly in the introduction, while democratic leaders can enhance commitment credibility through audience costs and electoral accountability mechanisms (Fearon 1994; Schultz 1999), autocratic leaders are compelled to rely upon personalistic ties, clientelist relationships or demonstrated resolve in order to render their commitments credible (Svolik 2012). In the context of entrapment risk, these facts suggest that China’s ability to credibly signal its concerns to North Korea – and North Korea’s incentives to heed those concerns – may operate differently than in democratic alliance contexts. North Korea’s leadership does not face electoral punishment for entrapping China, though it may face other costs such as reduced Chinese economic support or even abandonment. In terms of the massive power asymmetry that exists between the allies, Morrow’s (1991) security-autonomy trade-off model predicts that asymmetric alliances involve exchanges wherein larger allies provide security while smaller allies surrender policy autonomy. This suggests that China may possess greater leverage over

projected alliance outcomes that matter and not, as in the first argument, (intrinsically valuable) ally rights and obligations. The fact that the two arguments take different paths to reach the same normative conclusion – i.e. that the Moderate View is correct – is not problematic in that the two paths do not contradict, but rather complement each other. Thus the Moderate View can be correct because it is appropriately sensitive to the respective rights and obligations of the allies towards one another *and* because it leads to better ally- and alliance-level consequences than does a competing interpretation of if and when the CDD might justifiably be breached.

North Korean behaviour than would exist in a symmetric alliance. However, this leverage is not unlimited, particularly when the smaller ally possesses a source of bargaining leverage, such as the ability to engage in reckless behaviour that threatens its larger ally's interests. Such recklessness is, of course, precisely the kind of behaviour that entrapment entails (Benson 2012; Lanoszka 2018). The critical point here is not that regime type or power asymmetry can somehow nullify entrapment concerns, but rather that these alliance features affect how entrapment manifests itself in specific alliance cases and situations, as well as how the Moderate View's conditions (1) and (2) should be interpreted. With these qualifications in mind, let us turn to the hypothetical scenario.¹⁹

Let us imagine that North Korea desires to entrap China in a conflict with North Korea's eternal rival, South Korea. North Korea's objective in bringing about this conflict, which (as in the France-Greece case) it plans to insidiously provoke South Korea into initiating, is the forcible unification of the Korean peninsula under the authoritarian rule of the current North Korean regime. The North Korean leadership is aware that China maintains mutually beneficial economic relations and fairly good political relations with South Korea; that China opposes the unification of the Korean peninsula at present due to the presumed chaos this would cause and even more strongly opposes such unification by force; and that any attack on South Korea, even if disguised as self-defensive, would trigger the participation of South Korea's ally the United States, along with perhaps Japan, thereby threatening escalation to a region-wide, and potentially nuclear, great power war that would run roughshod over core Chinese interests. Put otherwise, the North Korean leadership is aware that it would be entrapping China if it embarked on such a perilous strategy of militarised unification. China, for its part, has reasonably deduced the existence of the objective, as well as its high likelihood and imminence, from a combination of public statements by the North Korean leadership, private statements by North Korean officials and diplomats

¹⁹ The worry might be raised here that the Moderate View is less applicable to such alliances because autocratic leaders are less constrained by moral obligations in the sense of more easily able to breach international commitments without domestic political costs. This worry misunderstands the purpose of the Moderate View. That purpose is the articulation of conditions under which CDD breach is morally justified. The Moderate Claim neither entails nor assumes that all states will either recognise or adhere to its normative terms and prescriptions. The empirical fact, then, that China's autocratic regime faces weaker domestic accountability for alliance decisions does not change the normative fact that, at least according to the Moderate View, China would be morally justified in breaching its CDD given imminent North Korean entrapment. What *may* differ between democratic and autocratic alliance contexts is the likelihood that moral prescriptions like those flowing from the Moderate View will *actually constrain* state behaviour. But this is obviously an empirical question concerning potential variation in state compliance with a moral prescription. It does not concern the normative applicability of that prescription as such.

to their Chinese counterparts, and intelligence gathered through the Chinese embassy in Pyeongyang. The Chinese premier is dispatched to Pyeongyang to unambiguously convey China's misgivings to the North Korean leadership. North Korea listens but dismisses these misgivings as exaggerated. Finally, the Chinese president himself places a phone call to the North Korean leader. Instead of merely reiterating Chinese entrapment concerns, the president explains in detail China's normative interpretation of the CDD as part of its official doctrine on collective defence ethics. This interpretation turns out to be synonymous in substance with the Moderate View, and the Chinese president subsequently hints at China's justified breach of its CDD obligations were North Korea to entrap it in a needless, irresponsible and *de facto* offensive war with South Korea and the United States for domination of the Korean peninsula. In essence, the Chinese president insinuates, should it carry out this entrapment strategy, North Korea would be fighting alone not only this time, but on future occasions as well, unless and until it forswore such entrapment.

What should we expect North Korea to do in this situation, assuming that it is at least a minimally rational actor that privileges above all its own continued survival, which would be put at grave risk were it to go to war with South Korea and the United States while China stayed neutral? Arguably, we have good reason to expect North Korea, once made aware of the Moderate View and its status as official Chinese doctrine, to decline to implement its planned policy of entrapment. Furthermore, assuming that North Korea values the considerable security and economic benefits furnished by its alliance with China, we also have good reason to expect North Korea to think very hard about the expected costs of engaging in entrapping behaviour in the future. And if we have good reason to expect these actions by North Korea, then we are entitled to at least tentatively infer that, all else being equal, the risk of entrapment should be lower in an alliance whose members accept that (1) a high likelihood of one member's entrapment and (2) an insensitivity to that member's signalled concerns, taken together, justify CDD violation by that member. More succinctly, the anti-entrapment and other harmony- and cohesion-promoting consequences (for the alliance) of subscribing to the Moderate View can themselves count as reasons in favour of that view. The China–North Korea case illustrates how the Moderate View operates in the context of an asymmetric, authoritarian alliance.

Objections to the Moderate View

In this section, I briefly consider and then reject three hypothetical objections to the Moderate View. I call these objections, respectively: the 'all-alliances-entrap' objection; the 'uncertainty-of-entrapment' objection; and the 'insufficient-harm' objection. First, the all-alliances-entrap objection. This objection criticises the Moderate View on the grounds that, allegedly, *all* alliances involve entrap-

ment by definition. Indeed, according to this objection, entrapment is one of the central functions of alliances, since by allying states formally commit to doing something that otherwise they would likely not do and that even currently they may have little strategic interest in doing – militarily defending another state. Thus entrapment is the very *point* of an alliance. Unfortunately, this objection confuses entrapment with the closely related but substantively distinct concept of ‘entanglement’. Tongfi Kim nicely distinguishes between these two concepts:

[Entanglement is] the process whereby a state is compelled to aid an ally in a costly and unprofitable enterprise *because of the alliance*. Entrapment is a form of *undesirable* entanglement [that occurs when an ally] adopts a *risky or offensive* policy not specified in the alliance agreement. (2011: 355, emphasis in original)

Kim goes on to stress that: ‘In order for states to benefit from alliances, they have to accept some risk of entanglement, because the benefits come from the possibility of entanglement. However, states can in fact benefit from alliances without accepting the risk of entrapment’ (2011: 355). Kim’s widely accepted distinction suggests that the all-alliances-entrap objection is misguided. All alliances cannot be said to entrap; instead, all alliances can be said to entangle, with entrapment constituting a perverse and injurious subtype of entanglement. The distinction is vital: whereas entanglement is a (necessary) built-in cost of any decision to ally, entrapment is a (contingent) defect or pathology of a specific alliance, a perverse and injurious subtype of entanglement.²⁰ As a result, the objection might be relabelled, somewhat tautologically, the ‘all-alliances-*entangle*’ objection; yet this new formulation would not apply to the Moderate View, as the latter conditions justified breach of the CDD on entrapment, which is avoidable, and not entanglement, which is not. Therefore, if the Moderate View as a normative argument is going to fail, it is not going to be because of the logical inevitability of allied entrapment.

Next, the uncertainty-of-entrapment objection. This objection attacks the Moderate View by pointing to the inherent fallibility of Ally X’s belief that Ally Y is planning to entrap it in the near future. Because that belief might very well be erroneous in regard to both the imminence of the entrapment and the high likelihood thereof, the Moderate View is wrong in thinking that it could ever justify CDD breach. This objection is surely correct about the fallibility inherent to making *ex ante* estimates of the probability of future entrapment. That being said, the objection is overdrawn, as it conflates a reason to be conscientious and cautious about future estimates of entrapment imminence and likelihood with

²⁰ See Beckley (2015), however, for an argument that even entanglement can be a ‘myth’.

a reason to deny the feasibility of such estimation outright. One need not, either logically or empirically, go as far as bowing to the latter imperative to be warranted in accepting the former imperative. To see why not, it is worth invoking, by analogy, the legality and ethics of preemptive versus preventive war. A preemptive war is staged in response to an epistemically near-certain and temporally imminent attack. By contrast, a preventive war is one in which both the degree of epistemic confidence in the attack's materialisation is weaker *and* the temporal proximity of the attack is more distant compared to a war staged preemptively. Crucially, both types of wars are founded on the fallible *ex ante* beliefs of the 'preempting' or 'preventing' state. Nevertheless, genuinely preemptive wars are generally considered legal under international law as well as morally permissible by most just war theorists. Preventive wars, by contrast, are almost universally considered illegal and, by most just war theorists, morally prohibited.²¹ If international law and just war theory can independently make this normative distinction between preemptive and preventive war, then the Moderate View's justification of CDD breach in the face of imminent and highly likely entrapment cannot be rejected just because future entrapment is always, technically, uncertain and beliefs in its imminence always, technically, defeasible. A principled distinction can thus be drawn between the Moderate View's *preemptive* prescription and a genuinely *preventive* prescription of CDD breach, with the latter justified by non-imminent and moderately probable entrapment. This distinction blocks much of the force of the uncertainty-of-entrapment objection.

Lastly, the insufficient-harm objection. This objection contends that the Moderate View fails because the imminent and highly probable threat of entrapment is insufficiently harmful to Ally X's security interests to justify breach of the CDD. The threat of entrapment may harm Ally X's security interests in an absolute sense; yet, once the harm to be caused to Ally Y by Ally X's breach of the CDD is factored in, the latter response becomes disproportionate and so unjustified. This objection is mistaken for two reasons. First, it does not take seriously enough the significant harm that can accompany being entrapped by an ally. Recall the two empirical illustrations above: It is very counterintuitive that, in these illustrations, French and Chinese security interests are not being seriously harmed by entrapment. France faces the prospect of what would likely be a lengthy and brutal war against a major military power – Turkey; a war that has the potential to engulf the entire Mediterranean region (and perhaps beyond), severely jeopardising French military assets, economic investments and political influence. China faces the prospect, not only of a war against a major

²¹ There are, however, (qualified) exceptions in just war theory: see Buchanan and Keohane (2004); Luban (2004).

regional military (South Korea's), but also of escalation to a great power conflict between itself and the United States. Such a conflict would, at a minimum, profoundly destabilise the interdependent security and economic architectures that have been in place in the region since the end of the Second World War and that, from 1978 to the present day, have directly facilitated China's rise to great power status. At a maximum, such a great power war could threaten the institutional and normative foundations of the entire postwar international order from which China, all things considered, continues to benefit. Second, the objection does not take seriously enough the extraordinary gravity of being responsible for entrapping an ally. Returning to the empirical illustrations, it is obvious that Greece and North Korea, if they implemented their entrapment plans, would be blameworthy for causing the considerable harm to their respective allies just discussed. This blameworthiness for such large-scale harm would plausibly render Greece and North Korea liable to suffer the retributive harm of CDD breach. Moreover, this breach would be temporary and open to reversal, unlike the irreversibility of entrapment itself. Consequently, imposing it on Greece and North Korea, far from being disproportionate, seems if anything an act of restraint. The insufficient-harm objection, then, is unpersuasive as a repudiation of the Moderate View.

Conclusion

For states, one of the greatest risks of participating in a military alliance is that of being left high and dry by one's ally in the face of an armed attack. In this article, I have adopted a normative approach to the question of if and when an ally can be justified in breaching what I have called its collective defence duty – i.e. its duty to defend its ally in the event that the latter suffers an attack. After rejecting two extreme views on this question, I have offered what I have termed the Moderate View. The Moderate View draws on Glenn Snyder's influential concept of the alliance security dilemma in the service of proposing that a state can be justified in breaching its collective defence obligations vis-à-vis an ally if two conditions hold. First, the state must have a reasonable belief that it faces a high likelihood of imminent entrapment by its ally. Second, the state must not be able to reduce this likelihood by signalling its entrapment concerns to its ally. I have put forward two distinct arguments in favour of the Moderate View, one grounded in the rights and obligations possessed by allies in relation to one another, the other grounded in the expected anti-entrapment consequences of allied acceptance of the Moderate View. Finally, I have defended the Moderate View against several potential criticisms.



Funding

No funding information.

Conflict of interest statement

No conflict of interest was reported.

Data availability statement

No supplementary data associated with the article.

Acknowledgements

I would like to thank the members of the CEJISS Editorial Team as well as the article's two anonymous reviewers for the substantial improvements to the article that were made possible by their excellent comments and vigorous critiques.

DAVID RUBIN is currently an independent researcher. He holds a PhD in Politics from the University of Manchester (UK). His primary research interest lies in analyzing the normative dimensions and implications of a variety of international political phenomena, such as system structure, armed conflict, collective defense, and audience costs.

References

Adler, E. (2008): The Spread of Security Communities: Communities of Practice, Self-Restraint, and NATO's Post-Cold War Transformation. *European Journal of International Relations*, 14(2), 195–230.

Adler, E. & Barnett, M. (eds.) (1998): *Security Communities*. Cambridge: Cambridge University Press.

Albert, E. (2019): The China-North Korea Relationship. *Council on Foreign Relations*, <accessed online: <https://www.cfr.org/backgrounder/china-north-korea-relationship>>.

Barnett, M. N. (1996): Identity and Alliances in the Middle East. In: Katzenstein, P. J. (ed.): *The Culture of National Security: Norms and Identity in World Politics*. New York: Columbia University Press, 327–364.

Benson, B. V. (2012): *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. Cambridge: Cambridge University Press.

Buchanan, A. & Keohane, R. O. (2004): The Preventive Use of Force: A Cosmopolitan Institutional Proposal. *Ethics and International Affairs*, 18(1), 1–22.

Deyermond, R. (2018): The Collective Security Treaty Organization. In: Tsygankov, A. P. (ed.): *Routledge Handbook of Russian Foreign Policy*. London: Routledge, 421–429.

Fearon, J. D. (1994): Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review*, 88(3), 577–592.

Gaubatz, K. T. (1996): Democratic States and Commitment in International Relations. *International Organization*, 50(1), 109–139.

Hemmer, C. & Katzenstein, P. J. (2002): Why is There No NATO in Asia? Collective Identity, Regionalism, and the Origins of Multilateralism. *International Organization*, 56(3), 575–607.

Jervis, R. (1978): Cooperation Under the Security Dilemma. *World Politics*, 30(2), 167–214.

Johnston, A. I. (2001): Treating International Institutions as Social Environments. *International Studies Quarterly*, 45(4), 487–515.

Jonah, A. (2025): Article 5: Trump Reopens Debate on NATO's Mutual Defence Pledge. *France 24*, <accessed online: <https://www.france24.com/en/europe/20250625-article-five-donald-trump-reopens-debate-nato-mutual-defence-pledge-usa>>.

Kant, I. (1993): *Grounding for the Metaphysics of Morals* (Third Edition). Translated by Ellington, J. W. New York: Hackett Publishing.

Kim, T. (2011): Why Alliances Entangle But Seldom Entrap States. *Security Studies*, 20(3), 350–377.

Krivosheev, K. (2021): Does the Collective Security Treaty Organization Have a Future? *Carnegie Moscow Center*, <accessed online: <https://carnegiemoscow.org/commentary/84923>>.

Kunertova, D. (2017): One Measure Cannot Trump It All: Lessons From NATO's Early Burden-Sharing Debates. *European Security*, 26(2), 552–574.

Lanoszka, A. (2018): Military Alliances in the Twenty-First Century. *Polity*, 50(1), 24–41.

Lazar, S. (2020): War. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*, Spring 2020 Edition, <accessed online: <https://plato.stanford.edu/entries/war/>>.

Leeds, B. A., Ritter, J. M., Mitchell, S. M. & Long, A. G. (2002): Alliance Treaty Obligations and Provisions, 1815–1944. *International Interactions*, 28(3), 237–260.

Luban, D. (2004): Preventive War. *Philosophy and Public Affairs*, 32(3), 207–248.

March, J. G. & Olsen, J. P. (1998): The Institutional Dynamics of International Political Orders. *International Organization*, 52(4), 943–969.

McMahan, J. (2009): *Killing in War*. Oxford: Oxford University Press.

Mearsheimer, J. J. (2001): *The Tragedy of Great Power Politics*. New York: Norton.

Morgenthau, H. J. (2006): *Politics Among Nations: The Struggle for Power and Peace*, Seventh Edition. Boston: McGraw-Hill.

Morrow, J. D. (1991): Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances. *American Journal of Political Science*, 35(4), 904–933.

North Atlantic Treaty (1949): *The Avalon Project*, <accessed online: https://avalon.law.yale.edu/20th_century/nato.asp>.

Pattison, J. (2012): *Humanitarian Intervention and the Responsibility to Protect: Who Should Intervene?* Oxford: Oxford University Press.

Pressman, J. (2008): *Warring Friends: Alliance Restraint in International Politics.* Ithaca: Cornell University Press.

Psaropoulos, J. (2021): Greece Ratifies Landmark Intra-NATO Defense Pact with France. *Al Jazeera News*, <accessed online: <https://www.aljazeera.com/news/2021/10/7/greece-ratifies-intra-nato-defense-pact-with-france>>.

Rawls, J. (1971): *A Theory of Justice.* Cambridge: Harvard University Press.

Rawls, J. (1999): *The Law of Peoples.* Cambridge: Harvard University Press.

Risse, T. (1996): Collective Identity in a Democratic Community: The Case of NATO. In: Katzenstein, P. J. (ed.): *The Culture of National Security: Norms and Identity in World Politics.* New York: Columbia University Press, 292–326.

Roehrig, T. & Heo, U. (2018): *The Evolution of the South Korea-United States Alliance.* Cambridge: Cambridge University Press.

Rubin, D. (2023): Transcending Two Percent: Toward a Prioritarian Model of NATO Burden-Sharing. *Central European Journal of International and Security Studies*, 17(4), 38–62.

Sandler, T. (2006): *Regional Public Goods: Typologies, Provision, Financing, and Development Assistance.* Stockholm: Almqvist & Wiksell International.

Schultz, K. A. (1999): Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War. *International Organization*, 53(2), 233–266.

Smith, A. (1995): Alliance Formation and War. *International Studies Quarterly*, 39(4), 405–425.

Snyder, G. H. (1984): The Security Dilemma in Alliance Politics. *World Politics*, 36(4), 461–495.

Snyder, G. H. (1997): *Alliance Politics.* Ithaca: Cornell University Press.

Sokou, K. (2021): Greece and France Give European Strategic Autonomy a Shot. *Atlantic Council*, <accessed online: <https://www.atlanticcouncil.org/blogs/new-atlanticist/greece-and-france-give-european-strategic-autonomy-a-shot/>>.

Stitt, B. G. & James, G. G. (1984): Entrapment and the Entrapment Defense: Dilemmas for a Democratic Society. *Law and Philosophy*, 3(1), 111–132.

Svolik, M. W. (2012): *The Politics of Authoritarian Rule.* Cambridge: Cambridge University Press.

Thaler, M. (2018): Unhinged Frames: Assessing Thought Experiments in Normative Political Theory. *British Journal of Political Science*, 48(4), 1119–1141.

Tomz, M. (2007): Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization*, 61(4), 821–840.

Tomz, M. & Weeks, J. L. P. (2021): Military Alliances and Public Support for War. *International Studies Quarterly*, 65(3), 811–824.

Vienna Convention on the Law of Treaties (1969): *United Nations Office of Legal Affairs*, <accessed online: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf>.

Vu, K. (2021): Why China and North Korea Decided to Renew a 60-Year-Old Treaty. *Lowy Institute*, <accessed online: <https://www.lowyinstitute.org/the-interpreter/why-china-north-korea-decided-renew-60-year-old-treaty>>.

Walt, S. M. (1987): *The Origins of Alliances*. Ithaca: Cornell University Press.

Waltz, K. N. (1979): *Theory of International Politics*. Reading: Addison-Wesley.

Walzer, M. (2015): *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 5th ed. New York: Basic Books.

Weeks, J. L. (2008): Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization*, 62(1), 35–64.

Weeks, J. L. (2012): Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict. *American Political Science Review*, 106(2), 326–347.

Wenar, L. (2020): Rights. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*, Spring 2021 Edition, <accessed online: <https://plato.stanford.edu/entries/rights/>>.

Wendt, A. (1999): *Social Theory of International Politics*. Cambridge: Cambridge University Press.

Wendt, A. (2004): The State as Person in International Theory. *Review of International Studies*, 30(2), 289–316.