European Journal of International Security (2024), page 1 of 18
doi:10.1017/eis.2024.14

EJIS

RESEARCH ARTICLE

Can you have it both ways? Attribution and plausible deniability in unclaimed coercion

Costantino Pischedda[1], Andrew Cheon[2] and Sara B. Moller[3]

[1]Department of Political Science, University of Miami, Coral Gables, FL, USA; [2]Duke Kunshan University, Kunshan, China and [3]Security Studies Program SSP, Georgetown University, Washington, DC, USA
Corresponding author: Costantino Pischedda; Email: costantino.pischedda@gmail.com

(Received 10 January 2023; revised 6 March 2024; accepted 12 March 2024)

Abstract

States and non-state actors conduct unclaimed coercive attacks, inflicting costs on adversaries to signal resolve to prevail in a dispute while refraining from claiming or denying responsibility. Analysts argue that targets often know who is responsible, which enables coercive communication, and that the lack of claims of responsibility grants coercers plausible deniability in the eyes of third parties. The puzzle of different audiences holding different beliefs about who is behind an unclaimed attack, even when they may have the same information, has been neglected. We address this puzzle by theorising that targets and third parties tend to reach different conclusions due to distinct emotional reactions: targets are more likely to experience anger, which induces certainty and a desire to blame someone, as well as heuristic and biased information processing, prompting confident attribution despite the limited evidence. A vignette-based experiment depicting a terrorist attack lends empirical plausibility to our argument.

Keywords: anger; appraisal tendencies; attribution; coercion; plausible deniability; unclaimed attacks

In 2000, a boat packed with explosives rammed the Navy destroyer USS Cole in the port of Aden, Yemen, killing 19 US sailors. The US government suspected that al Qaeda, which had previously vowed to expel American forces from Muslim lands, was behind the unclaimed attack.[1] Nonetheless, both Presidents Bill Clinton and George W. Bush decided not to retaliate, due to the absence of evidence to 'nail down responsibility with certainty'.[2] In more recent times, Israel reportedly carried out a series of unclaimed attacks against Iranian assets in Syria to discourage Tehran from hostile actions.[3] As these examples show, both states and non-state actors engage in unclaimed coercive acts, that is, they inflict costs on adversaries to signal their resolve to prevail in a dispute while refraining from claiming, or denying, responsibility. Unclaimed terrorist and cyber attacks employed for coercive purposes, as well as instances of coercive covert action, fall under this rubric.[4]

[1] 9/11 Commission, 'Final Report of the National Commission on Terrorist Attacks upon the United States' (22 July 2004), pp. 190–214, available at: {https://9-11commission.gov/report/}.

[2] Michael Morrell, *The Great War of Our Time: The CIA's Fight Against Terrorism. From Al Qa'ida to ISIS* (New York: Twelve, 2015), p. 38.

[3] Robin Wright, 'Israel wages a growing war in Syria', *The New Yorker* (10 April 2018), available at: {https://www.newyorker.com/news/news-desk/israel-wages-a-growing-war-in-syria}.

[4] This section and the following draw on Costantino Pischedda and Andrew Cheon, 'Does plausible deniability work? Assessing the effectiveness of unclaimed coercive acts in the Ukraine war', *Contemporary Security Policy*, 44:3 (2023), pp. 345–71. By covert action we mean 'a variety of secret foreign policy actions that may be administered by military or intelligence bureaucracies … in a way that conceals and renders deniable the role of the sponsoring state for most audiences'. Austin

Scholars and commentators often note that the targets of unclaimed coercive acts know who is behind them, allowing the coercive message to be understood.[5] Existing analyses also tend to posit that the lack of a claim of responsibility provides actors with the benefit of plausible deniability: given that it may not be possible to definitely pin the action on them, perpetrators may be shielded from retaliation by the target and third parties, as well as from international opprobrium.[6] The study of unclaimed coercive acts has not fully appreciated the puzzle deriving from the juxtaposition of these two common observations: how could targets of unclaimed acts identify perpetrators with enough confidence for the intended coercive message to go through, while sufficient doubt remains in the minds of third parties to protect perpetrators from the negative repercussions of their actions? How could anonymous coercers have it both ways?

Recent works on covert action advance an important answer: the target of coercion has privileged access to information about the identity of the perpetrator but refrains from disclosing it publicly, either to prevent the escalation of the dispute or to avoid compromising intelligence sources and future collection opportunities.[7] While this perspective sheds light on a number of unclaimed coercive acts, it cannot explain instances in which targets are all too eager to share intelligence implicating the perpetrator, such as when Kyiv blew Russia's cover on its eastern Ukraine intervention with photographic evidence.[8] In fact, states often express concerns about, and go to considerable lengths to avoid, exposure of their covert operations by adversaries and targets.[9]

Drawing on the literature on the appraisal tendencies of emotions, we theorise a different solution to the puzzle. We posit that, independent of the information at their disposal, targets and third parties tend to form different attribution beliefs about unclaimed attacks due to their respective emotional reactions. Specifically, targets tend to experience anger, an emotion that leads individuals to feel a sense of certainty and a desire to blame someone.[10] Moreover, while the anger-induced feeling of certainty works as internal evidence that one's understanding of the situation is accurate, thus obviating the need for further information search, anger also makes individuals prone to dismiss evidence at odds with their beliefs.[11]

As a result of these psychological processes, we expect targets to display a tendency to confidently attribute unclaimed attacks to a plausible culprit despite objective gaps in the evidence. By contrast, third parties should be less likely to experience anger, making them less inclined to

Carson and Keren Yarhi-Milo, 'Covert communication: The intelligibility and credibility of signaling in secret', *Security Studies*, 26:1 (2017), pp. 124–56 (p. 128).

[5] See, for example, Daniel Byman and Sarah Kreps, 'Agents of destruction? Applying principal-agent analysis to state-sponsored terrorism', *International Studies Perspective*, 11:1 (2010), pp. 1–18; and Dennis Pluchinsky, 'The terrorism puzzle: Missing pieces and no boxcover', *Terrorism and Political Violence*, 9:1 (1997), pp. 7–10.

[6] See, for example, Rory Cormac and Richard Aldrich, 'Grey is the new black: Covert action and implausible deniability', *International Affairs*, 94:3 (2018), pp. 477–94; Klaas Voß, 'Plausibly deniable: Mercenaries in US covert interventions during the Cold War, 1964–1987', *Cold War History*, 16:1 (2016), pp. 37–60.

[7] Allison Carnegie and Austin Carson, *Secrets in Global Governance: Disclosure Dilemmas and the Challenge of International Cooperation* (Cambridge: Cambridge University Press, 2020); Austin Carson, *Secret Wars: Covert Conflict in International Politics* (Princeton, NJ: Princeton University Press, 2018); Carson and Yarhi-Milo, 'Covert communication', pp. 124–56.

[8] Andrew Higgins, Michael R. Gordon, and Andrew E. Kramer, 'Photos link masked men in East Ukraine to Russia', *New York Times* (20 April 2014), available at: {https://www.nytimes.com/2014/04/21/world/europe/photos-link-masked-men-in-east-ukraine-to-russia.html}.

[9] Michael Joseph and Michael Poznansky, 'Media technology, covert action, and the politics of exposure', *Journal of Peace Research*, 55:3 (2018), pp. 320–35.

[10] See, in particular, Jennifer. S. Lerner and Larissa. Z. Tiedens, 'Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition', *Journal of Behavioral Decision Making*, 19:2 (2006), pp. 115–37.

[11] Jennifer S. Lerner, Julie H. Goldberg, and Philip E. Tetlock, 'Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility', *Personality and Social Psychology Bulletin*, 24:6 (1998), pp. 563–74; Michael MacKuen, Jennifer Wolak, Luke Keele, and George E. Marcus, 'Civic engagements: Resolute partisanship or reflective deliberation', *American Journal of Political Science*, 54:2 (2010), pp. 440–58.

attribute unclaimed attacks with confidence in the face of limited evidence. To use Mercer's terminology, attribution of anonymous attacks is an 'emotional belief', resulting from a confluence of cognition and emotion.[12]

We probe the plausibility of the argument with a vignette experiment. We randomise whether respondents were told that their country or some other country was the target of a major terrorist attack that, according to intelligence agencies and the media, was probably carried out by a (fictional) foreign armed group, though it had not claimed responsibility. Subjects were then asked to identify the perpetrator of the unclaimed attack, with a range of possible degrees of confidence, and to report their own feelings. Crucially, this empirical approach allows us to hold constant respondents' information and prior beliefs, thus making sure that different attribution tendencies are a function of the experimental manipulation.

Consistent with our theory of anger-driven attribution, individuals from the target country are statistically and substantially more likely to confidently attribute unclaimed attacks than individuals from another country. Mediation analysis reveals anger to be one of the mechanisms for the observed difference in attribution tendencies, indicating that individuals from the target country tend to attribute unclaimed attacks more frequently because they are angry.

Shedding light on the attribution of unclaimed coercive acts is important because attribution is likely to shape targets' responses. Attribution is generally necessary for coercive success, as the identity of the coercer is often inextricably tied to demands: a target's inferences about concessions being tacitly demanded with an unclaimed act would differ depending on whether, say, Russia or Iran is thought to be behind the act. Similarly, retaliation requires a 'return address', that is, a clearly identified wrongdoer that can be punished. Furthermore, retaliation against an actor wrongly believed to be behind an unclaimed attack has the potential of unleashing a process of inadvertent escalation.

The article makes three specific contributions to our collective understanding of attribution. First, it reconciles the prevalent view about the intelligibility of unclaimed acts as coercive messages with the commonly held notion that the absence of a claim of responsibility grants perpetrators the benefit of plausible deniability, thus providing a theoretically coherent answer to an underexplored puzzle. To wit, we show how a logic of strategic interaction that appears to work in practice can work in theory. Second, the article contributes to a growing body of research on the influence of emotions on coercive bargaining. Much of this research focuses on how anger shapes the prospects for coercive success and escalation. Our contribution is distinctive, as we focus on how anger affects the attribution of unclaimed attacks. Third, our argument about anger-driven attribution suggests unexplored pathways for emotions' influence on spiral dynamics.[13] Angry attribution may accentuate cognitive biases, which make individuals impervious to evidence indicating that an adversary is less aggressive than previously thought and too quick to dismiss the possibility that an event may have resulted from an accident rather than being part of an adversary's hostile design, thus contributing to escalatory processes.

The puzzle of unclaimed coercion

Coercive bargaining sometimes takes an unclaimed form, with perpetrators refraining from claiming, or denying, responsibility for their actions. For example, while many observers suspect Russia to have launched a series of unclaimed cyberattacks in 2007 to extract concessions from Estonia, Moscow denied involvement.[14] The unclaimed 1988 bombing of Pan Am Flight 103 over Lockerbie is now widely thought to have been orchestrated by Muammar Gaddafi's government in retaliation against earlier US air strikes on Libya.[15]

[12] Jonathan Mercer, 'Emotional beliefs', *International Organization*, 64:1 (2010), pp. 1–31.

[13] Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976).

[14] Brandon Valeriano, Benjamin Jensen, and Ryan Maness, *Cyber Strategy: The Evolving Character of Cyber Power and Coercion* (Oxford: Oxford University Press, 2018), pp. 124–7.

[15] Bruce Hoffman, 'Why terrorists don't claim credit', *Terrorism and Political Violence*, 9:1 (1997), pp. 1–6 (p. 4).

Unclaimed coercive actions extend beyond inter-state relations. For instance, the Tamil Tigers reportedly conducted attacks on both military and civilian targets to induce the Sri Lankan government to concede Tamil self-determination, but they refrained from claiming attacks against civilians.[16] In fact, the majority of acts of terrorism, often seen as the most typical example of coercion, go unclaimed.[17] This is not to say that all (or even most) unclaimed actions by states and non-state actors are coercive in nature. Some seek information (e.g. espionage), on-the-ground effects (e.g. sabotaging an adversary's military capabilities), or strategic surprise – goals more likely to be attained if the action itself remains secret.[18] However, the above examples suggest that an empirically important share of unclaimed actions is coercive, warranting our focus on understanding how different audiences attribute them.

Coercive bargaining entails threatening to inflict (and/or actually inflicting) costs on adversaries to send a message about one's resolve to prevail in a dispute, such as 'we are willing to take any necessary step to prevent you from getting what you want' or 'we will keep causing you pain until you concede'. Coercers attempt to shape targets' calculus by convincing them that the costs of not complying outweigh the benefits. Coercive bargaining, therefore, is an act of communication, requiring the transmission of an intelligible message from coercers to targets.[19] Coercive success would be nearly impossible if targets were completely in the dark about the identity of perpetrators of coercive acts, as targets might not know what is being demanded of them.

The fact that a coercive action is unclaimed, however, does not necessarily imply utter uncertainty about who is behind it, as the list of plausible culprits – those with both motives and capabilities – is often short. In fact, commentators and scholars envision targets as typically having a clear sense of who their anonymous tormentor is, which enables the coercive message to go through. At the same time, commentators and scholars often see the absence of a claim of responsibility as providing perpetrators with the benefit of plausible deniability.[20] Absent a claim of responsibility and airtight evidence about the identity of the culprit, the unclaimed act cannot be conclusively attributed to the actor, thus shielding it from possible retaliation by the target and third parties, as well as from international opprobrium.[21]

[16] Pluchinsky, 'The terrorism puzzle', p. 4.

[17] Vincent Bauer, Keven Ruby, and Robert Pape, 'Solving the problem of unattributed political violence', *Journal of Conflict Resolution*, 61:7 (2017), pp. 1437–564; Martha Crenshaw and Gary LaFree, *Countering Terrorism* (Washington, DC: Brookings, 2017), pp. 131–64; Erin M. Kearns, 'When to take credit for terrorism? A cross-national examination of claims and attributions', *Terrorism and Political Violence*, 33:1 (2021), pp. 164–93; Keir Lieber and Daryl Press, 'Why states won't give nuclear weapons to terrorists', *International Security*, 38:1 (2013), 80–104.

[18] In other cases of unclaimed acts that are not coercive, the public nature of the act, but not its attribution to a given actor, is necessary, given that these actions do not seek specific concessions from targets. For instance, terrorist attacks may be part of a spoiling strategy, aiming to erode trust in a peace process, or a destabilisation strategy, seeking to create a climate of chaos. Erin M. Kearns, Brendan Conlon, and Joseph K. Young, 'Lying about terrorism', *Studies in Conflict & Terrorism*, 37:5 (2014), pp. 422–39; Andrew H. Kydd and Barbara F. Walter, 'The strategies of terrorism', *International Security*, 31:1 (2006), pp. 49–80. On secrecy and strategic surprise in inter-state war, see Branislav L. Slantchev, 'Feigning weakness', *International Organization*, 64:3 (2010), pp. 357–88.

[19] Thomas Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966).

[20] Andrew Bowen, 'Coercive diplomacy and the Donbas: Explaining Russian strategy in eastern Ukraine', *Journal of Strategic Studies*, 42:3–4 (2019), pp. 312–43; Cormac and Aldrich, 'Grey is the new black'; Joseph and Poznansky, 'Media technology, covert action, and the politics of exposure'; Voß, 'Plausibly deniable'. Poznansky distinguishes between the 'state model' of plausible deniability, aiming to obfuscate the involvement of the state in a covert operation, and the 'executive model', aiming to shield chief executives from responsibility. As most International Relations scholarship on plausible deniability, this article focuses on the former, though extending its application to non-state actors as perpetrators. Michael Poznansky, 'Revisiting plausible deniability', *Journal of Strategic Studies*, 45:4 (2020), pp. 511–33.

[21] With enough time, investigators may be able to gather sufficient evidence to convince third parties about the culpability of the suspect. However, retaliating long after an attack may be practically or politically unfeasible. See Crenshaw and LaFree, *Countering Terrorism*, pp. 150–8; Cormac and Aldrich, 'Grey is the new black', p. 480. Plausible deniability may also enable governments to avoid domestic political costs. See, for example, Alexander B. Downes and Mary L. Lilley, 'Overt peace, covert war? Covert intervention and the democratic peace', *Security Studies*, 19:2 (2010), pp. 266–306.

Though rarely stated this bluntly, Pluchinsky's observation about unclaimed terrorist attacks conveys the essence of this influential perspective on unclaimed coercion: 'The target knows why and who but cannot prove it', which in turn discourages retaliation.[22] For example, Byman and Kreps posit that Iran's covert reliance on the services of Hezbollah furthers Tehran's coercive diplomacy against Israel and the United States because their retaliation would be politically complicated in the absence of smoking-gun evidence of Iranian involvement.[23] Analogously, in a discussion of the alleged Iran-sponsored attack on Saudi oil facilities in 2019, Carson notes that 'neither the United States nor the Saudis would be able to justify a harsh military response' if the 'attack can't be definitively pinned on Iran. … Ambiguity would make any response look illegitimate while jeopardizing allies' support.'[24] Thus, from Tehran's point of view, 'the ideal outcome could be leaving enough evidence of its involvement to send a signal of strength to rivals while leaving enough ambiguity to take the teeth out of any response'.[25]

An unexplored puzzle underlies the perspectives just outlined: how could targets confidently attribute unclaimed attacks while third parties remain uncertain about attribution? What explains the fact that these different actors hold diverging beliefs about the identity of the perpetrator?

Recent studies point to asymmetric information as the answer. Carson and Yarhi-Milo argue that bargaining through unclaimed coercive acts is possible because 'the basic contours of covert behavior are often visible' to the parties to the dispute but not necessarily to others, which makes private, intelligible communication possible.[26] Adversaries may collude to prevent the exposure of one another's covert operations to other audiences, as keeping the dispute to the 'backstage' helps contain international reputation and domestic audience costs as well as escalation risks.[27] For example, the United States concealed intelligence indicating that Soviet pilots were covertly fighting in the Korean War and thus engaged in a secret, limited war against the Soviet Union. Yet adversarial relations in which the target publicly identifies the perpetrator of unclaimed coercive action – instead of colluding by staying silent – seem to abound. In the cases of anonymous attacks mentioned at the beginning of this section, Estonia, the United States, and Sri Lanka openly pointed their fingers at Russia, Libya, and the Tamil Tigers, respectively.

Asymmetric information could be the key to the puzzle even in the absence of collusion between adversaries. Targets' confident attribution could result from their access to intelligence implicating the suspect, which may not be shareable with third parties due to concerns about jeopardising sources and future collection. Without access to this intelligence, however, third parties would be unable to verify targets' claims, which may thus lack credibility – a predicament that Carnegie and Carson dub the 'disclosure dilemma'.[28]

Though constraints on intelligence sharing represent an important explanation for disagreements about who is behind an unclaimed attack in some cases, in other cases diverging attribution beliefs appear to hinge on different interpretations of the *same* information. For instance, given the high levels of intelligence cooperation within the alliance, constraints on intelligence sharing are an unlikely explanation for the fact that, while Estonian leaders accused Russia of carrying out a wave of cyberattacks in 2007 against the Baltic country, officials from other NATO members

[22] Pluchinsky, 'The terrorism puzzle', p. 8. Some analysts of cyber warfare are less sanguine about the ease of attribution by the target of unclaimed attacks. Yet these analysts too note that, in cases in which the target is relatively confident about 'who did it', not enough evidence may be available to convince third parties, complicating retaliation. See, in particular, Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica, CA: RAND Corporation, 2009), pp. 41–2.

[23] Byman and Kreps, 'Agents of destruction?', pp. 4–6.

[24] Austin Carson, 'After the Saudi oil attack, will the U.S. and Saudis start a war with Iran? Here are 3 things to know', *Washington Post* (17 September 2019), available at: {https://www.washingtonpost.com/politics/2019/09/17/after-saudi-oil-attack-will-us-saudis-start-war-with-iran-here-are-things-know/}.

[25] Ibid.

[26] Carson and Yarhi-Milo, 'Covert communication', p. 132.

[27] Carson, *Secret Wars*.

[28] Carnegie and Carson, *Secrets in Global Governance*.

refrained from attribution.[29] The 2019 unclaimed attacks against oil tankers in the Gulf of Oman represent another relevant example. The attacks occurred after Tehran threatened to close the Strait of Hormuz, the main export outlet for Middle Eastern oil, in response to Washington's efforts to ramp up restrictions on the flow of Iranian oil. The target of Iranian coercive diplomacy – the US government – claimed Tehran was responsible for the attacks, while third parties, such as Japan and Germany, argued that the available evidence was insufficient for attribution.[30] While the possibility that differences in attribution were driven by access to different information cannot be ruled out, the twin facts that the Trump administration engaged in significant diplomatic efforts to persuade Japan and Germany of Iran's responsibility and that the two countries are close US allies cast serious doubt on constraints to intelligence sharing as an explanation in this case.

To be clear, conflicting attribution statements may not always reflect genuine differences in beliefs. In particular, states may misrepresent the extent to which they are confident about the identity of the perpetrator of an unclaimed attack to hurt a rival or help an ally. For example, it is no coincidence that the only countries to have attributed the unclaimed attack on the Kakhovka dam in Ukraine to Kyiv are Syria and Belarus, two allies of Russia, while a number of backers of Ukraine accused Moscow.[31] However, the fact that the four main providers of aid to Ukraine – the United States, Germany, the United Kingdom, and France – refrained from claiming that Russia was responsible also suggests that statements about attribution are not simply reducible to geopolitical interests.

In sum, we do not doubt that in some cases asymmetries of information and interests play an important role in explaining divergent assessments about who is responsible for unclaimed attacks reached by targets and third parties. However, as the examples above suggest, these asymmetries are likely not the only factors at play. Thus, we put forth a new explanation centred around the effects of the different actors' emotional responses on their attribution beliefs.

Anger-driven attribution of unclaimed attacks

Laypersons, philosophers, and social scientists alike traditionally saw emotion as clouding rational judgement and decision-making. Neuroscience research over the past few decades, however, has debunked the notion that emotion is antithetical to rationality, showing that emotion provides a foundation for swift and accurate judgement and decision-making and is thus necessary for rationality.[32] Without emotion, we would be like medical patients with abnormal or damaged parts of the brain critical to processing emotional information, retaining normal cognitive skills yet incapable of steering away from self-destructive life decisions or of reaching any decision at all.[33] Though emotion can compromise rationality, rationality requires emotion.

Anger is one of the most studied emotions in International Relations and International Security. Much of the literature focuses on the 'action tendencies' of anger, that is, how actors tend to respond

[29] Stephen Herzog, 'Revisiting the Estonian cyber attacks: Digital threats and multinational responses', *Journal of Strategic Security*, 4:2 (2011), pp. 49–60; Ian Traynor, 'Russia accused of unleashing cyberwar to disable Estonia', *Guardian* (16 May 2007), available at: {https://www.theguardian.com/world/2007/may/17/topstories3.russia}.

[30] Carol Morello, Kareem Fahim, and Simon Denyer, 'Standoff with Iran exposes Trump's credibility issue as some allies seek more proof of tanker attack', *Washington Post* (16 June 2019), available at: {https://www.washingtonpost.com/world/saudi-crown-prince-blames-iran-for-tanker-attacks-as-tensions-soar/2019/06/16/7eeb43ca-900c-11e9-b162-8f6f41ec3c04_story.html}.

[31] Alonso Gurmendi, 'Tracking state reactions to the destruction of the Kakhovka dam', Opinion Juris blog (20 June 2022), available at: {http://opiniojuris.org/2023/06/20/tracking-state-reactions-to-the-destruction-of-the-kakhovka-dam/}.

[32] Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: G.P. Putnam, 1994); Rose McDermott, 'The feeling of rationality: The meaning of neuroscientific advances for political science', *Perspective on Politics*, 2:4 (2004), pp. 691–706; Mercer, 'Emotional beliefs'.

[33] Antoine Bechara, Daniel Tranel, and Hanna Damasio, 'Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions', *Brain*, 123:11 (2000), pp. 2189–202; Damasio, *Descartes' Error*; Antonio Verdejo-García, Jose M. Pérez-García, and Antoine Bechara, 'Emotion, decision-making and substance dependence: A somatic-marker model of addiction', *Current Neuropharmacology*, 4:1 (2006), pp. 17–31.

to angering events. For example, Hall argues that displays of anger can help states signal escalation risks and that this emotion can provoke states into self-destructive courses of action;[34] McDermott, Lopez, and Hatemi theorise about how the potential for retaliation in anger can strengthen the credibility of nuclear deterrent threats in settings where using nuclear weapons would be suicidal due to an adversary's second strike capability;[35] and Snyder investigates how human rights violators' angry reactions may cause efforts to induce compliance through naming-and-shaming to backfire.[36] In contrast to this literature, we focus on the 'appraisal tendencies' of anger – the ways in which this emotion affects what and how people think – to theorise about the effects of anger on attribution of unclaimed attacks.[37] We leave in abeyance the important question of the effects of anger-induced attribution on targets' behavioural responses to unclaimed coercive acts.

Numerous experimental studies indicate that anger induces a general sense of certainty and a tendency to assign blame for the angering event to someone who could then be targeted in retaliation.[38] The feeling of certainty induced by anger tends to trigger heuristic processing, that is, superficial and simplified examination of information – feeling certain operates as internal evidence that one's understanding of the situation is already accurate and no further processing is necessary.[39] Furthermore, anger has a distinct tendency to produce motivated biases in information processing, favouring information that supports one's views and dismissing contrary evidence.[40] Other emotions have different appraisal tendencies. Fear, for instance, induces a sense of uncertainty and prompts systematic processing of information. Importantly, existing theory and evidence suggest that we should not expect anger to have an effect on judgement in the absence of ambiguous events serving as inkblots open to interpretation.[41]

We argue that these appraisal tendencies of anger – sense of certainty and tendency to assign blame, as well as heuristic and biased information processing – spur confident attribution of

[34] Todd Hall, 'We will not swallow this bitter fruit: Theorizing a diplomacy of anger', *Security Studies*, 20:4 (2011), pp. 521–55; Todd Hall, 'On provocation: Outrage, International Relations, and the Franco–Prussian War', *Security Studies*, 26:1 (2017), pp. 1–29.

[35] Rose McDermott, Anthony Lopez, and Peter Hatemi, 'Blunt not the heart, enrage it: The psychology of revenge and deterrence', *Texas National Security Review*, 1:1 (2017), pp. 68–88.

[36] Jack Snyder, 'Backlash against human rights shaming: Emotions in groups', *International Theory*, 12:1 (2020), pp. 109–32.

[37] Markwica's book represents an important exception, as it studies both appraisal and action tendencies of emotions, including anger, in coercive diplomacy. Robin Markwica, *Emotional Choices: How the Logic of Affect Shapes Coercive Diplomacy* (Oxford: Oxford University Press, 2018).

[38] Mark D. Alicke, 'Culpable control and the psychology of blame', *Psychological Bulletin*, 126:4 (2000), pp. 556–74; Julie H. Goldberg, Jennifer S. Lerner, and Philip E. Tetlock, 'Rage and reason: The psychology of the intuitive prosecutor', *European Journal of Social Psychology*, 29:5–6 (1999), pp. 781–95; Jennifer S. Lerner and Dacher Keltner, 'Fear, anger, and risk', *Journal of Personality and Social Psychology*, 81:1 (2001), pp. 146–59; Lerner and Tiedens, 'Portrait of the angry decision maker'. Other appraisal tendencies of anger that are less relevant to attribution of unclaimed attacks are a sense of high individual (as opposed to situational) control and optimism.

[39] Galen V. Bodenhausen, Lori A. Sheppard, and Geoffrey P. Kramer, 'Negative affect and social perception: The differential impact of anger and sadness', *European Journal of Social Psychology*, 24:1 (1994), pp. 45–62; Lerner, Goldberg, and Tetlock, 'Sober second thought'; Larissa Z. Tiedens and Susan Linton, 'Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing', *Journal of Personality and Social Psychology*, 81:6 (2001), pp. 973–88.

[40] MacKuen, Wolak, Keele, and Marcus, 'Civic engagements'; Elizabeth Suhay and Cengiz Erisen, 'The role of anger in the biased assimilation of political information', *Political Psychology*, 39:4 (2018), pp. 793–810.

[41] Lerner and Keltner, 'Fear, anger, and risk'. Experimental studies of the guilt phase of criminal trials illustrate the effect of anger's appraisal tendencies on attribution in contexts in which there is a suspect, but the evidence of culpability is ambiguous. For example, Bright and Goodman-Delahunty report that mock jurors are more likely to find defendants guilty and be more confident about the sufficiency of the prosecution's evidence when shown gruesome crime pictures that make them angry. David. A. Bright and Jane Goodman-Delahunty, 'Gruesome evidence and emotion: Anger, blame, and jury decision-making', *Law and Human Behavior*, 30:2 (2006), pp. 183–202. See also Susan Bandes and Jessica Salerno, 'Emotion, proof and prejudice: The cognitive science of gruesome photos and victim impact statements', *Arizona State Law Journal*, 46:4 (2014), pp. 1003–56; David. A. Bright and Jane Goodman-Delahunty, 'The influence of gruesome verbal evidence on mock juror verdicts', *Psychiatry, Psychology and Law*, 11:1 (2004), pp. 154–66; and Kevin S. Douglas, David R. Lyon, and James R. P. Ogloff, 'The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: Probative or prejudicial?', *Law and Human Behavior*, 21:5 (1997), pp. 485–501.

unclaimed attacks by their targets to a plausible culprit, but less angry third parties remain relatively uncertain about the identity of the perpetrator. Therefore, unclaimed attacks may indeed enable perpetrators to send intelligible coercive messages to targets while enjoying the benefits of plausible deniability before other audiences.

Targets are more likely to be angry, due to their direct exposure to coercion, than third parties. Anger generates a desire to blame someone for the angering event, thus directing targets' attention towards a plausible culprit. The anger-induced sense of certainty makes the available evidence appear more inculpatory and/or lowers subjective standards of proof for attribution, while heuristic and biased processing prompts angry targets to pay limited attention to or dismiss information pointing in a different direction. Put simply, targets' anger fills the evidentiary gaps about unclaimed attacks. By contrast, in cases of claimed attacks we should not expect the attribution tendencies of targets and third parties to differ: claims of responsibility largely remove the ambiguity of the situation, thus suppressing anger's influence on targets' judgement.

It should be noted that the appraisal tendencies of anger may not spur confident attribution if there is no plausible culprit in the first place. However, in practice this is not a highly restrictive condition. Typically, there is a shortlist of plausible perpetrators, with a particular state or armed group standing out based on a mix of capabilities and motives. Yet though intelligence and forensic work can further chip away at the uncertainty, in the absence of a smoking gun or a credible confession, attribution entails an element of subjective judgement about the probative value of the available evidence and the appropriate standard of proof. How do various pieces of evidence, by themselves and in combination, affect the probability that a given actor is responsible? How much evidence is enough for confident attribution? The ambiguity of the situation provides room for targets' anger to influence their judgement, prompting them to identify someone to blame and boosting confidence in the attribution beyond the available evidence.[42]

Our argument and alternative explanations for differences in attribution of unclaimed attacks centred around asymmetries of information and interests are not mutually exclusive. We expect anger to cause a divergence between the attribution beliefs of targets and third parties independent of differences in the information at their disposal or in their interests; yet evidence that these differences also affect attribution beliefs would not contradict our argument. Assessing the relative explanatory power of anger-driven attribution and competing explanations is a task for future empirical research. Nonetheless, our argument has the advantage of being relevant to a broad range of cases, including those without substantial asymmetries of information and interests between targets and third parties.

We draw three hypotheses from our argument:

H1: *Targets will be more likely to confidently attribute unclaimed attacks than third parties.*

H2: *Targets will be more likely to experience anger than third parties.*

H3: *Anger will mediate the effect of being a target, as opposed to a third party, on confident attribution of unclaimed attacks.*

[42] From our theoretical perspective, targets and third parties can be likened to family members of homicide victims and jurors in the corresponding criminal trials, respectively. Bereaved family members tend to be more confident in the culpability of the defendant and to disagree with juries' acquittal or exoneration verdicts. Though multiple psychological processes are at play, anger induced by the loss of loved ones is probably a key factor behind these diverging attribution judgements. Sarah Goodrum, 'Bridging the gap between prosecutors' cases and victims' biographies in the criminal justice system through shared emotions', *Law & Social Inquiry*, 38:2 (2013), pp. 257–87 (pp. 273–4); Samuel R. Gross and Daniel J. Matheson, 'What they say at the end: Capital victims' families and the press', *Cornell Law Review*, 88:2 (2003), pp. 486–516 (pp. 507–10); Darren Thiel, 'Moral truth and compounded trauma: The effects of acquittal of homicide defendants on the families of the victims', *Homicide Studies*, 20:3 (2016), pp. 199–219 (pp. 214–16).

Coercion could succeed directly, by persuading the government of the target country that the expected costs of defiance are not worth the expected benefits, or indirectly, by prompting the population of the target country to pressure its government into making concessions.[43] Thus, both the government and the population of the country under coercive attack should be considered as 'targets'. In light of the above-mentioned findings from neuroscience about the fundamental role played by emotions in rational judgement and decision-making and the specific evidence that the beliefs of foreign policy elites, like those of other humans, are shaped by emotions, we do not expect a systematic discrepancy in the attribution of unclaimed attacks by these two audiences.[44] In other words, our argument applies to both foreign policy elites and ordinary citizens.

Empirical approach

Disentangling the effect of being the target, rather than a third party, on attribution of anonymous attacks from other causes represents a significant inferential challenge. Targets and third parties may not only have access to different information about an anonymous attack but may also have different incentive structures and priors about suspected perpetrators, which could affect attribution. To sidestep these concerns, we probe the plausibility of our argument with an experimental approach, rather than examining historical cases of unclaimed coercion. In particular, we use a vignette-based experiment manipulating whether subjects are from the target country or another country, while holding constant possible confounders such as respondents' information, incentives, and priors.

The vignettes refer to fictional events and actors to minimise the possibility of subjects relying on their attitudes and priors about real-world actors in their answers.[45] The prompt informed subjects that they would read about a fictional scenario but noted that comparable events had occurred in the past and may occur again.

The analytical drawback is that our fictional scenario is unlikely to induce intense emotional reactions in the subjects. The use of a nominal, rather than ordinal, scale to measure emotions – asking subjects to select the term from a list that best describes how they feel – is thus helpful for our plausibility probe: correctly identifying one's main emotional reaction entails a lower, more realistic bar to clear for subjects' emotional intelligence than providing an accurate assessment on a five-point scale of various possible reactions experienced at low intensity.[46]

The low intensity of subjects' emotional responses points to concerns about ecological validity, which apply more broadly to the experimental study of the effects of emotion in contexts of intense political conflict: due to obvious practical and ethical constraints, researchers cannot induce emotional responses even remotely approaching the intensity of those caused by real-world episodes of coercion and political violence.[47] We believe this fact makes observational studies of the political effects of emotions important complements of experiments, as each approach helps overcome

[43] Robert A. Pape, *Bombing to Win: Air Power and Coercion in War* (Ithaca, NY: Cornell University Press, 1996).

[44] Jonathan Mercer, 'Emotion and strategy in the Korean War', *International Organization*, 67:2 (2013), pp. 221–52.

[45] Examples of studies using vignette experiments with fictional scenarios and hypothetical actors include articles by Brutger and Kertzer and by Tomz. Ryan Brutger and Joshua D. Kertzer, 'A dispositional theory of reputation costs', *International Organization*, 72:3 (2018), pp. 693–724; Michael Tomz, 'Domestic audience costs in international relations: An experimental approach', *International Organization*, 61:4 (2007), pp. 821–40.

[46] On nominal and ordinal approaches to measuring emotions, see Klaus R. Scherer, 'What are emotions? And how can they be measured?', *Social Science Information*, 44:4 (2005), pp. 695–729 (p. 717). A possible problem with an ordinal scale is that subjects might score two emotional responses (say, fear and anger) at the lowest level on the scale because they are experiencing them with much less intensity than in the context of real-life events, even though subjects actually feel one of the two marginally more intensively than the other. Another possibility is a variant of the central tendency bias: as subjects probably would not feel strongly either one of those two emotional responses, they might score both at an intermediate level on the scale, without making the introspective effort of assessing whether there is in fact a difference in intensity between the two.

[47] Roger D. Petersen, *Western Intervention in the Balkans: The Strategic Use of Emotion in Conflict* (Cambridge: Cambridge University Press, 2011), p. 30.

the other's limitations.[48] Yet the general challenge of the ecological validity of the experimental study of emotions in contexts of intense political conflict has the benefit of making our specific research design relatively demanding for the argument we put forth: if even the relatively weak treatment of an unambiguously fictional terrorist attack can arouse enough anger to affect attribution judgements, then we should be reasonably confident that the more intense and long-lasting anger unleashed by a real-world terrorist attack would have a meaningful effect on attribution outside the experimental setting.[49]

Our sample comprises 402 adults residing in the United States, recruited through the Qualtrics online survey platform with sampling quotas to match US Census statistics for gender, age, and education.[50] Since, as noted, our hypotheses apply to both the population and the government of the target country, ideally we would conduct survey experiments with samples of the general population and foreign policy elites. However, given the substantial practical obstacles to conducting survey experiments with foreign policy elites, we employ a general population sample to obtain insights at both levels of analysis. This approach is warranted for two reasons.

First, using a general population sample may offer indirect insight into policymakers' responses. In fact, a number of studies indicate that, despite differences in expertise and backgrounds, foreign policy elites' responses to experimental manipulation and real-world events resemble those of the general public.[51] Second, public reactions may operate as constraints on foreign policy. For example, policymakers risk significant losses in public support if they do not stand firm in the face of foreign provocation, a particularly serious concern in democracies.[52]

The vignette reported the occurrence of a terrorist attack that caused hundreds of casualties and contained a colour photograph of the physical destruction, depicting a building torn apart by an explosion and surrounded by rubble. Our argument about the differential attribution tendencies of targets and third parties due to anger should apply to a broad range of unclaimed attacks, regardless of type of perpetrator (state or non-state actor), the modality of attack (cyber or kinetic), and the status of people and objects (civilian or military) directly affected. For this initial probe, we decided to focus on a major terrorist attack by a non-state actor under the assumption that it would appear plausible to our US-based respondents and it would have more potential of inducing

[48] See, for example, Stuart J. Kaufman, *Nationalist Passions* (Ithaca, NY: Cornell University Press, 2015); Markwica, *Emotional Choices*; Mercer, 'Emotion and strategy in the Korean War'.

[49] For instance, anger was the most commonly experienced emotion by US citizens in response to the 11 September 2001 attacks, both in the immediate aftermath and after a year. See Mitja D. Back, Albrecht C. P. Küfner, and Boris Egloff, 'The emotional timeline of September 11, 2001', *Psychological Science*, 21:10 (2010), pp. 1417–19; Jennifer S. Lerner, Roxana M. Gonzalez, Deborah A. Small, and Baruch Fischhoff, 'Emotion and perceived risks of terrorism: A national field experiment', *Psychological Science*, 14:2 (2003), pp. 144–50; Baruch Fischhoff, Roxana M. Gonzales, Jennifer S. Lerner, and Deborah A. Small, 'Evolving judgments of terror risks: Foresight, hindsight, and emotion', *Journal of Experimental Psychology: Applied*, 11:2 (2005), pp. 124–39.

[50] The experiment was registered on 1 November 2019 and conducted in mid-December 2019. We calculated the required sample size based on the objective of detecting with 80 per cent power a 20 per cent difference in the rate of confident attribution of attacks between subjects from another country (50 per cent) and subjects from the target country (70 per cent). Thus, a minimum of 186 subjects are needed to study attribution of unclaimed attacks and another 186 for claimed attacks, for a total of 372. We rounded up our sample to 400 (the additional two subjects were already taking the survey when the 400 quota was met). The research was approved by the Johns Hopkins University's Homewood Institutional Review Board.

[51] Hall, 'On provocation'; Joshua D. Kertzer, Jonathan Renshon, and Keren Yarhi-Milo, 'How do observers assess resolve?', *British Journal of Political Science*, 51:1 (2021), pp. 308–30; Joshua D. Kertzer, 'Re-assessing elite–public gaps in political behavior', *American Journal of Political Science*, 66:3 (2022), pp. 539–53; Mercer, 'Emotion and strategy in the Korean War'; Keren Yarhi-Milo, *Who Fights for Reputation: The Psychology of Leaders in International Conflict* (Princeton, NJ: Princeton University Press, 2018); Keren Yarhi-Milo, Joshua D. Kertzer, and Jonathan Renshon, 'Tying hands, sinking costs, and leader attributes', *Journal of Conflict Resolution*, 62:10 (2018), pp. 2150–79.

[52] Allan Dafoe, Samuel Liu, Brian O'Keefe, and Jessica Chen Weiss, 'Provocation, public opinion, and international disputes: Evidence from China', *International Studies Quarterly*, 66:2 (2022), pp. 1–14; Alexander Debs and Jessica Chen Weiss, 'Circumstances, domestic audiences, and reputational incentives in international crisis bargaining', *Journal of Conflict Resolution*, 60:3 (2016), pp. 403–33; Michael Tomz, Jessica L. P. Weeks, and Keren Yarhi-Milo, 'Public opinion and decisions about military force in democracies', *International Organization*, 74:1 (2020), pp. 119–43.

anger than cyberattacks and covert acts by adversaries, which are typically less dramatic and lethal for civilians.[53] Studying the effects of different types of unclaimed attacks represents an important step for future research.

We adopted a between-subjects design, randomising two aspects of the vignette, while holding everything else constant: (1) whether the target of the attack was the respondent's own country ('your country') or another country ('a foreign country'), and (2) whether the attack was unclaimed or claimed, captured by the dummy variables TARGET and UNCLAIMED, respectively.[54] In both vignettes with claimed and unclaimed attacks, subjects were informed that intelligence and media reports pointed to the 'Sons of Freedom' (SOF), a fictional terrorist organisation seeking the withdrawal of all foreign troops from its homeland 'Ruritania', as the likely culprit. The unclaimed attack vignette noted that the group did not claim responsibility, while the claimed attack vignette reported that SOF had claimed the attack. Thus, each of the following four vignettes had about 100 respondents: (a) the subject's own country is the target of a claimed attack; (b) the subject's own country is the target of an unclaimed attack; (c) a claimed attack targets a country other than the respondent's; (d) an unclaimed attack targets a country other than that of the respondent.[55] The 'your country' vs. 'other country' manipulation allows us to examine our hypotheses, while the claimed vs. unclaimed manipulation enables an assessment of our argument's scope condition that anger should not affect attribution of claimed attacks in the absence of substantial ambiguity.

After reading their randomly assigned vignette, all subjects were asked which actor they thought had carried out the attack – the list included Sons of Freedom, an unspecified other terrorist organisation, the country from which Sons of Freedom hailed, the target country, and an unspecified other country – and how confident they were in that assessment – 'not confident at all', 'somewhat confident', and 'very confident'. We used these answers to create a measure of confident attribution, ATTRIBUTION, taking on 1 when respondents reported being at least somewhat confident that SOF had conducted the attack, and 0 otherwise.

Subjects were then asked which one of the following terms better described how they felt: 'fear', 'indifference', 'resignation', 'confusion', 'anger', or 'other'. (Respondents who answered 'other' were asked to type in how they felt.) We relegated subjects' self-reporting of feelings to the end of the survey, as labelling one's emotions has been shown to reduce their impact on judgement.[56] Together with anger, fear is a key emotion in studies of coercive bargaining in a variety of contexts, whether discussed in explicit or implicit emotional terms.[57] Fear has distinct appraisal tendencies: low certainty and systematic (as opposed to heuristic) processing.[58] While individuals experiencing fear focus their attention on the potential threat to select the most appropriate response (fight, flight, or freeze), the associated feeling of low certainty prompts efforts to process information carefully and in depth instead of jumping to conclusions. Thus, unlike anger, fear should not mediate the positive effect of being from the target country on the probability of confident attribution of an unclaimed attack.

[53] Highly dramatic and lethal kinetic covert actions by states are, of course, possible (e.g. the Lockerbie bombing by Libya), but they have not directly affected the United States in recent years, a fact that might have reduced the plausibility of a scenario depicting such an event in the eyes of our respondents.

[54] The online appendix reports the vignettes and the corresponding questions (available, with other replication materials, at https://doi.org/10.7910/DVN/BLBF53).

[55] Before seeing the vignette, subjects were asked a battery of general questions, which we used for robustness checks.

[56] Dacher Keltner, Kenneth D. Locke, and Paul C. Audrain, 'The influence of attributions on the relevance of negative feelings to personal satisfaction', *Personality and Social Psychology Bulletin*, 19:1 (1993), pp. 21–9.

[57] See, for example, Alexander L. George, *Forceful Persuasion: Coercive Diplomacy as an Alternative to War* (Washington, DC: US Institute of Peace, 1991); Kydd and Walter, 'The strategies of terrorism'; Lerner, Gonzalez, Small, and Fischhoff, 'Emotion and perceived risks of terrorism'; Petersen, *Western Intervention in the Balkans*.

[58] Lerner and Keltner, 'Fear, anger, and risk'; Tiedens and Linton, 'Judgment under emotional certainty and uncertainty'. Fear has other appraisal tendencies that are less relevant to attribution of unclaimed attacks, in particular, a sense of situational (as opposed to individual) control and pessimism.

We included resignation and indifference in our list of possible feelings to offer subjects the option of reporting subdued emotional responses (with a negative valence in the case of resignation). Moreover, we listed confusion, given that this feeling is sometimes noted as an effect (whether intended or unintended) of unclaimed attacks.[59] The literature does not suggest that the appraisal tendencies of these last three reactions – resignation, indifference, confusion – should mediate the positive effect of TARGET on ATTRIBUTION. In fact, it seems plausible that confusion, commonly thought of as a feeling of uncertainty and/or limited understanding, would have a negative effect on the probability of confident attribution.[60] A dummy variable corresponds to each of the self-reported feelings: ANGER, FEAR, RESIGNATION, CONFUSION, INDIFFERENCE, and OTHER.

A discrete emotions approach based on self-reporting is better suited for probing our argument than alternative approaches, increasingly used in political science, tracking physiological responses, such as skin-conductance reactivity, which can measure emotional arousal but cannot identify discrete emotions.[61] Although limits to subjects' ability to report on their emotional state and social desirability bias cannot be dismissed, two considerations suggest that these are not particularly serious concerns here. First, subjects were asked how they were feeling rather than how the event in the vignette made them feel, which should increase the probability that in their answers subjects would rely on introspection rather than on their implicit causal theories about how they *should* feel.[62] Second, the fact that our survey does not deal with particularly sensitive issues and is automatically administered online to anonymous subjects should curtail social desirability bias. Nonetheless, we recognise that to advance the understanding of anger-driven attribution beyond the plausibility probe the present study offers, future experimental research would benefit from also employing alternative approaches to both measurement and elicitation of emotional responses, such as coding participants' answers to open-ended questions about feelings being experienced and priming subjects with stimuli known to consistently induce anger (e.g. specifically validated movie clips).[63]

Analysis

According to our argument, targets of unclaimed attacks tend to confidently attribute them to a plausible culprit despite gaps in the evidence, but third parties remain relatively uncertain, thus enabling anonymous coercers to send an intelligible message to targets, while benefiting from

[59] See, for example, Aaron M. Hoffman, 'Voice and silence: Why groups take credit for acts of terror', *Journal of Peace Research*, 47:5 (2010), pp. 615–26 (p. 615). We sidestep the debate about whether confusion is an emotion, a mental state, or a metacognition. See Phoebe C. Ellsworth, 'Confusion, concentration, and other emotions of interest: Commentary on Rozin and Cohen (2003)', *Emotion*, 3:1 (2003), pp. 81–5; Ursula Hess, 'Now you see it, now you don't – The confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003)', *Emotion*, 3:1 (2003), pp. 76–80; Paul Rozin and Adam B. Cohen, 'High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans', *Emotion*, 3:1 (2003), pp. 68–75; Paul J. Silvia, 'Confusion and interest: The role of knowledge emotions in aesthetic experience', *Psychology of Aesthetics, Creativity, and the Arts*, 4:2 (2010), pp. 75–80.

[60] Of the 402 respondents, 26 chose the 'other' answer. Five of them typed in a description of their feelings compatible with dictionary definitions of one of the five listed answers – confusion (see the online appendix for a list of the 26 responses). Thus, we ran robustness checks coding these five subjects as expressing confusion. Our results are robust to this alternative coding and to dropping the 26 observations with 'other' as an answer (see Tables A25–A26 in the appendix).

[61] Jonathan Renshon, Julia Lee, and Dustin Tingley, 'Emotions and the micro-foundations of commitment problems', *International Organization*, 71:S1 (2017), pp. 189–218. Approaches to emotion measurement relying on behavioral indicators such as voice and facial expression are also generally ineffective at identifying discrete emotions. See Iris B. Mauss and Michael D. Robinson, 'Measures of emotion: A review', *Cognition and Emotion*, 23:2 (2009), pp. 221–6. Moreover, physiological and behavioral approaches require specialised equipment, which makes their use outside the laboratory impractical.

[62] Even Wilson, who has made a strong case for the 'adaptive unconscious', where a broad range of mental processes occur beyond individual awareness, acknowledges that cases 'in which people fail to recognize a feeling … may not be very common'. Timothy D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge, MA: Harvard University Press, 2004), p. 135.

[63] James K. Gross and Robert W. Levenson, 'Emotion elicitation using films', *Cognition and Emotion*, 9:1 (1995), pp. 87–108.

Fig. 1- B/W online, B/W in print

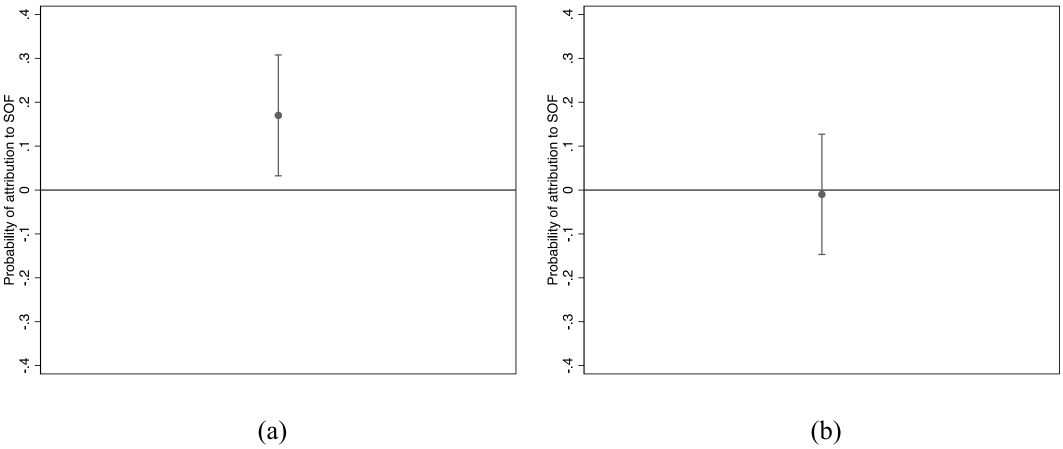

Figure 1. (a) Treatment effect of being from the target country on attribution of unclaimed attacks. (b) Treatment effect of being from the target country on attribution of claimed attacks.
Note: Figures 1a and 1b display the average treatment effects of a change in TARGET from 0 to 1 on the probability of ATTRIBUTION=1 (for unclaimed and claimed attacks, respectively) using OLS on the full sample of 402 subjects, with TARGET, UNCLAIMED, and their interaction as independent variables (regression output reported in Table A1 in the online appendix).

plausible deniability before other audiences. Thus, per H1, TARGET (i.e. the dummy indicating whether a respondent is from the target country rather than another country) should have a positive effect on the probability of ATTRIBUTION (i.e. confident attribution to SOF) of unclaimed attacks. By contrast, no such effect should be observed for claimed attacks, given that the presence of claim of responsibility dissipates much of the ambiguity surrounding the situation, thus setting these attacks outside the scope of the theory of anger-driven attribution.

Figure 1a reports the effects of the variable TARGET on ATTRIBUTION for unclaimed attacks. TARGET has the expected positive effect on attribution for unclaimed attacks. Respondents from the target country are 17 per cent more likely to confidently attribute an unclaimed attack to the plausible culprit than respondents from another country. This is a substantively large effect as it amounts to a 50 per cent increase in the share of respondents that confidently attribute the unclaimed attack, going from a clear minority (34%) to a majority (51%). As expected, TARGET displays no significant effect on ATTRIBUTION for claimed attacks (Figure 1b).

These results are robust to the inclusion of controls for the following pre-treatment variables (see Figures A1–A7 in the online appendix): gender, ideology, partisanship, level of national pride, age, education, and interest in current affairs.[64]

Our argument envisions anger as the underlying mechanism for the observed difference in attribution tendencies for unclaimed attacks. Individuals from the target country should be more likely to experience anger than individuals from another country, prompting the former to confidently attribute unclaimed attacks to SOF. As Table 1 indicates, a plurality of subjects from the target country – 33 per cent – report anger as the best descriptor of their feelings, while the corresponding figure for subjects from another country is 20 per cent, which is lower than the share of those respondents reporting indifference. Table 2 displays the results of bivariate OLS regressions of subjects' being from the target country (TARGET) on self-reported feelings.

Consistent with H2, TARGET has a significant positive effect on ANGER; subjects from the target country are 13 per cent more likely to report experiencing anger than subjects from another country. Subjects from the target country are also more likely to experience fear, which makes

[64] Our treatment and control groups are well balanced in terms of pre-treatment variables, as reported in Table A2 in the online appendix. The only exception is the self-reported level of interest in current affairs, which is marginally higher for subjects reading about unclaimed attacks than those reading about claimed attacks.

Table 1. Descriptive statistics of self-reported dominant feeling.

Feeling	Respondents from target country (%)	Respondents from other country (%)
Anger	32.8	19.9
Fear	24.4	12.9
Resignation	7.5	10.9
Confusion	12.4	26.9
Indifference	17.9	21.4
Other	5	8

Table 2. Effect of being from target country (as opposed to another country) on reported feelings.

DV	Coefficient	SE
Anger	0.129***	0.044
Fear	0.114***	0.039
Resignation	−0.035	0.029
Confusion	−0.144***	0.039
Indifference	−0.035	0.040
Other	−0.030	0.025
Number of observations	402	

Note: Intercepts not reported. The independent variable is OWN COUNTRY.
Inference: *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

sense because an attack, albeit fictional, against one's own country should be perceived as a more of a threat to the self than an attack against some other country.

Similarly, the significant negative effect of CONFUSION on ATTRIBUTION makes sense in light of the appraisal tendency of certainty associated with anger: if subjects from the target country are more likely to experience a sense of certainty as a result of their feeling of anger, then they should be less likely to report confusion. By contrast, the absence of a significant effect for INDIFFERENCE is surprising, as we might have expected subjects from another country to be more likely to report indifference than subjects from the target country. This result might reflect social desirability bias, as some respondents may be concerned about appearing callous if they express their indifferent reaction to a terrorist attack, albeit a fictional one, against another country. Inasmuch as some of the subjects experiencing indifference to a terrorist attack targeting a foreign country opt to report a feeling of anger instead, this should result in an attenuation bias of the expected mediating effect of anger on attribution.

Respondents' personal interest in the events described in the vignettes represents a possible alternative explanation for our main finding that subjects from the target country and from another country display different tendencies to attribute unclaimed attacks. Though our experimental design does not provide incentives for attribution, subjects from the target country might perceive particularly high stakes, as they are accustomed to thinking of their personal security as connected to that of their own state. These subjects, therefore, may be more likely to engage in a substantial cognitive effort to attribute an unclaimed attack. The problem with this alternative explanation is that it is not clear why a more careful processing of the available information by subjects from the target country should result in more confident attribution; instead, it may well lead subjects to focus on the information about the lack of a claim of responsibility and thus prompt agnosticism rather than confidence about the identity of the perpetrator. By contrast, our argument provides a coherent explanation for the higher rate of attribution of unclaimed attacks among subjects from the target country: anger should prompt subjects to train their attention on the plausible culprit and to feel confident that the actor is responsible, in turn reducing attention to the contradictory

Fig. 2- B/W online, B/W in print

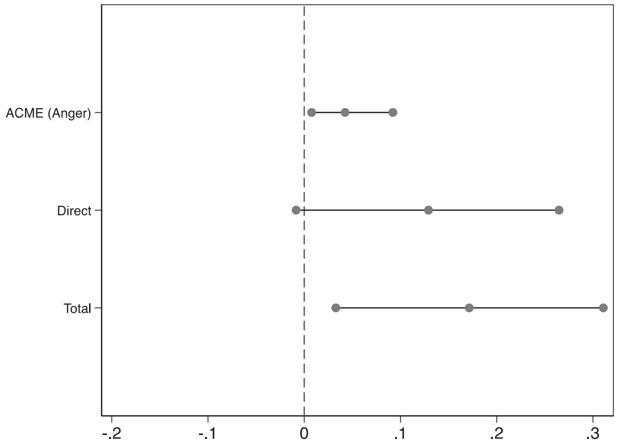

Figure 2. Anger as mediator for targets' attribution of unclaimed attacks.
The treatment is TARGET, the outcome is ATTRIBUTION, and the mediator is ANGER. The horizontal lines represent 95 per cent confidence intervals for the estimates. ACME is the average causal mediation effect of anger.

information about the lack of a claim of responsibility. Furthermore, if different levels of interest were behind our main finding, we might expect subjects from the target country to pay more attention to the task and thus spend more time reading the vignette and answering questions. However, there is no statistically significant difference in terms of survey duration between subjects from the target country and those from another country, and our results are robust to controlling for duration (see Table A3 and Figure A8 in the online appendix).

Next, we turn to an analysis of anger as a mediator of the effect of TARGET on ATTRIBUTION and thus to an evaluation of H3. Mediation analysis allows us to assess whether anger is a causal mechanism underlying the observed difference in attribution of anonymous attacks between targets and third parties, though it cannot rule out the existence of additional mechanisms. Following the procedure developed by Imai, Keele, and Tingley, we estimate the average causal mediation effect (ACME) of ANGER – the expected difference in the outcome when the mediator takes the value it would realise under the treatment condition (TARGET = 1) as opposed to the control condition (TARGET = 0), while the treatment is held fixed.[65] We also estimate the direct effect of the treatment, i.e. the expected difference in the outcome when TARGET goes from 0 to 1 while ANGER is held constant, and the total effect, that is, the sum of the direct and the mediation effects.

Figure 2 depicts the results of the mediation analysis. Consistent with H3, when the attack is unclaimed the ACME is positive and significant at the 95 per cent level, amounting to 25 per cent of the total treatment effect. (When the attack is claimed, no significant effect is observed; see Figure A10.) These results confirm that to the extent that the targeting of a subject's country induced anger, this prompted confident attribution to SOF of the unclaimed attack, though TARGET might be affecting ATTRIBUTION through other causal pathways as well.

Among the other self-reported feelings, only CONFUSION mediates the treatment effect (see Figures A11–A15 in the online appendix). The significant ACME for CONFUSION indicates that being from a country targeted by unclaimed attacks makes subjects less likely to feel confused, which in turn prompts them to confidently attribute the attack to SOF. We interpret this result as compatible with our argument. Being confused is by definition associated with a sense of uncertainty, so the reduced tendency for confused subjects to confidently attribute anonymous attacks is unsurprising; the lower likelihood of reporting confusion for subjects whose country has been

[65] Kosuke Imai, Luke Keele, and Dustin Tingley, 'A general approach to causal mediation analysis', *Psychological Methods*, 15:4, (2010), pp. 309–34. We use the STATA mediation package developed by Hicks and Tingley. Raymond Hicks and Dustin Tingley, 'Causal mediation analysis', *Stata Journal*, 11:4 (2012), 605–19.

targeted is consistent with the appraisal tendency of certainty associated with anger, an emotion that these subjects are more likely to report (the correlation between confusion and anger is −0.296). However, when both ANGER and CONFUSION are included in the same model using the approach for multiple mediators developed by Preacher and Hayes (2008), the indirect effect of CONFUSION loses statistical significance, while the indirect effect for ANGER remains significant (see Table A4).[66]

A potential concern for our mediation analysis is that an omitted variable may be influencing both the mediator and the outcome. For example, it could be that gender or political ideology affects anger and attribution. This would amount to a violation of the 'sequential ignorability assumption' on which mediation analysis relies.[67] Though we cannot rule out unobserved confounders, we assuage this concern by checking the robustness of our mediation findings to the inclusion of a range of pre-treatment variables – gender, ideology, partisanship, level of national pride, age, education, and interest in current affairs.[68] The effect of ANGER is robust throughout (Figures A16–A22 in online appendix).

In sum, we find support for our three hypotheses. Respondents from the target country are more likely both to experience anger and to confidently attribute anonymous attacks to plausible culprits. Mediation analysis indicates a causal connection between the two: individuals from the target country are more prone to confident attribution because of their anger.

Implications and directions for future research

Existing studies have not fully explored a pervasive puzzle in unclaimed coercive bargaining: how could targets of unclaimed attacks generally infer 'who did it', as many observers presume, while perpetrators enjoy the oft-noted benefits of plausible deniability in the eyes of third parties? How could the two sets of actors hold diverging attribution beliefs? We theorise that the different emotional reactions to unclaimed attacks of targets and third parties provide a key to the puzzle. Targets' direct exposure to wrongdoing triggers anger, which in turn activates a series of psychological dynamics leading to a higher probability of attributing an unclaimed attack to a plausible culprit, the objective evidentiary gaps notwithstanding. By contrast, third parties are less likely to experience anger and the corresponding influences on the content and processes of their thinking, resulting in a relatively low probability of confident attribution in the absence of a claim of responsibility.

Our experimental results indicate the empirical plausibility of the theoretical expectations that being the target, as opposed to a less directly affected observer, of an unclaimed attack increases the probability of confident attribution to a plausible culprit and that the target's anger mediates this effect. Future studies could probe and extend our initial findings in several ways. First, future studies could assess the ecological validity of our findings using a combination of cases studies of real-world instances of unclaimed coercion and survey experiments with samples of foreign policy elites. Second, while our approach holds subjects' priors and information constant, different experimental designs and careful process-tracing of policymakers' judgement and deliberation could be leveraged to assess the relative importance of anger and alternative (though compatible) arguments

[66] Kristopher Preacher and Andrew Hayes, 'Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models', *Behavioral Research Methods*, 40:3 (2008), pp. 879–91.

[67] Kosuke Imai, Luke Keele, and Teppei Yamamoto, 'Identification, inference and sensitivity analysis for causal mediation effects', *Statistical Science*, 25:1 (2010), pp. 51–71.

[68] This is a standard approach in the literature. See, for example, Renshon, Lee, and Tingley, 'Emotions and the microfoundations of commitment problems'. Unfortunately, we cannot conduct a sensitivity analysis as proposed by Imai, Keele, and Yamamoto, because that method cannot handle a set-up where both the mediator and the outcome variable are binary. See Imai, Keele, and Yamamoto, 'Identification, inference and sensitivity analysis'; Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto, 'Causal mediation analysis using R', Working paper (September 2019), available at {https://cran.ism.ac.jp/web/packages/mediation/vignettes/mediation-old.pdf}.

emphasising differences in information and incentives between targets and third parties as explanations for their divergent attribution beliefs. Third, future studies could assess the robustness of our findings by adopting alternative approaches to measuring and eliciting subjects' anger. A fourth avenue for further experimental research is exploring the extent to which anger affects attribution of unclaimed attacks that do not cause physical destruction, given that attribution and coercion in the cyber realm have been subjects of significant academic and policy debate.[69] Fifth, future studies should assess if our findings hold regardless of whether the perpetrators are state or non-state actors and of whether civilian or military objectives are affected. Sixth, future studies should also investigate the *effects* of anger-driven attribution of unclaimed coercive acts: does it tend to prompt targets to retaliate, comply, or do nothing? Under what circumstances does one type of response prevail?

As a body of research in neuroscience has shown, even though emotions can lead actors astray, they are necessary for rationality, rather than antithetical to it.[70] Similarly, our argument and findings that emotions affect attribution should not be interpreted as implying that irrational thinking dominates anonymous coercive bargaining. In the specific case of our vignette, we cannot say whether anger made accurate attribution more likely, as the identity of the perpetrator is left unspecified. Furthermore, we cannot tell whether targets' anger-driven attribution yields more false positives (cases in which a plausible culprit is wrongly accused by the target) than the false negatives resulting from third parties' reluctance to advance firm attribution judgements.[71] Therefore, our take-away message is that, in much the same way that emotions are involved in rational, strategic decisions such as launching preventive war, 'gambling for resurrection' to recoup serious losses, or trusting allies and rivals, anger plays an important role in the attribution process of unclaimed attacks.[72]

The key implication for both academics and policymakers is that approaches to coercive bargaining that ignore emotional beliefs are incomplete and thus potentially misleading. For example, though Gartzke's analysis has made important contributions in debunking the idea that the spread of cyber capabilities heralds a revolution in military affairs, its strict adherence to an emotionless logic of coercion reveals a blind spot.[73] The notion that the benefit of being shielded from retaliation provided by 'internet anonymity' comes at the huge cost of failing 'to provide the target with the means to acquiesce' to the coercer's demands makes perfect sense if emotions do not affect attribution.[74] However, in a world where 'feeling is believing', anonymous coercers may be well placed to successfully deploy cyber and traditional means of influence without sacrificing plausible deniability.[75] Similarly, Abrahms argues that the fact that most terrorist attacks go unclaimed discredits a dominant view of terrorism as following a coercion logic: terrorists must not be trying to extract

[69] See, for example, Erica Borghard and Shawn Lonergan, 'The logic of coercion in cyberspace', *Security Studies*, 26:3 (2017), pp. 452–81; Jon R. Lindsay and Erik Gartzke, 'Coercion through cyberspace: The stability–instability paradox revisited', in Kelly M. Greenhill and Peter J. Krause (eds), *Coercion: The Power to Hurt in International Politics* (New York: Oxford University Press, 2018), pp. 179–203; Travis Sharp, 'Theorizing cyber coercion: The 2014 North Korean operation against Sony', *Journal of Strategic Studies*, 40:7 (2017), pp. 898–926; and Valeriano, Jensen, and Maness, *Cyber Strategy*.

[70] Damasio, *Descartes' Error*.

[71] Thus, anger-driven attribution differs from the phenomenon of 'positive illusions' – the widespread tendency for individuals to be overconfident about their capabilities and prospects of success – which can be considered a deviation from rationality. See Dominic D. P. Johnson, *Overconfidence and War* (Cambridge, MA: Harvard University Press, 2004).

[72] Mercer, 'Emotional beliefs'.

[73] Erik Gartzke, 'The myth of cyberwar: Bringing war in cyberspace back down to earth', *International Security*, 38:2 (2013), pp. 41–73.

[74] Gartzke, 'The myth of cyberwar', p. 47.

[75] Gerald L. Clore and Karen Gasper, 'Feeling is believing: Some affective influences on belief', in Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem (eds), *Emotions and Beliefs: How Feelings Influence Thoughts* (Cambridge: Cambridge University Press, 2000), pp. 10–44.

concessions from states, given that anonymous attacks preclude targets from complying.[76] This article suggests that anonymous coercion is not an oxymoron. Though likely not aware of anger's effects on attribution of unclaimed attacks, terrorist organisations, like states, may be acting strategically under an intuitive understanding that somehow the targets will receive the intended message.

Finally, this article contributes to the integration of emotions in the spiral model. People's tendencies to explain others' behaviour in terms of dispositional qualities, rather than situational causes, and to discount the possibility that events may be the result of accidents rather than others' hostile plans increase the risk of actors' developing negative images of one another and of the intensification of conflict.[77] Anger could exacerbate these dynamics because it prompts individuals to look for a culprit to be punished even in circumstances in which a damaging event could have plausibly occurred by chance or at the initiative of a rogue agent, thus potentially bringing an unwarranted deterioration of relations between the two countries. Moreover, anger could strengthen the imperviousness of negative images about others to contrary evidence, which would make it difficult to interrupt an escalation of tensions and to promote reconciliation.[78] When angering events shape an actor's view of an adversary, updating in response to conciliatory gestures may become particularly unlikely, given anger's tendency to induce a sense of certainty and in turn to reduce attention to potentially inconsistent information.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eis.2024.14.

Acknowledgements. The authors would like to thank Jonthan Mercer, Jack Snyder, and the participants to the University of Miami's faculty colloquium for feedback on an early draft of the article. The authors gratefully acknowledge financial support from the University of Miami and Johns Hopkins University.

Dr Costantino Pischedda is Associate Professor in the Department of Political Science at the University of Miami. His research focuses on various aspects of international security, civil wars, and ethnic conflict. Dr Pischedda is the author of *Conflict among Rebels: Why Insurgent Groups Fight Each Other* (Columbia University Press) and articles published in *International Security, Security Studies, International Studies Quarterly, Journal of Conflict Resolution, Journal of Peace Research, Contemporary Security Policy*, and *Ethnopolitics*, among other journals.

Dr Andrew Cheon is Associate Professor of International Relations at Duke Kunshan University. His research focuses on pressing issues of governance, contestation, and conflict in the age of climate change and great power competition.

Dr Sara B. Moller is Associate Teaching Professor in the Security Studies Program (SSP) at Georgetown University and Non-Resident Senior Fellow at the Atlantic Council. Her current research examines organisational adaptation in alliances in peacetime and wartime. Her research has been published in the *Journal of Strategic Studies, Asian Security, International Politics, The Washington Quarterly, Survival*, and elsewhere.

[76] Max Abrahms, 'What terrorists really want: Terrorist motives and counterterrorism strategy', *International Security*, 32:4 (2008), pp. 78–105 (pp. 89–90).

[77] Edward E. Jones and Richard E. Nisbett, 'The actor and the observer: Divergent perceptions of the causes of behavior', in Edward E. Jones, David E. Kanouse, Harold H. Kelley, et al. (eds), *Attribution: Perceiving the Causes of Behavior* (Morristown, NJ: General Learning Press 1972), pp. 79–94; Jervis, *Perception and Misperception*, pp. 321–3.

[78] Jervis, *Perception and Misperception*, p. 68.

Cite this article: Costantino Pischedda, Andrew Cheon and Sara B. Moller, 'Can you have it both ways? Attribution and plausible deniability in unclaimed coercion', *European Journal of International Security* (2024), pp. 1–18. https://doi.org/10.1017/eis.2024.14