

RESEARCH ARTICLE

Sound the alarm! Updating beliefs and degradative cyber operations

Miguel Alberto N. Gomez*

Center for Security Studies at ETH Zurich

*Corresponding author. Email: miguel.gomez@sipo.gess.ethz.ch

(Received 13 June 2018; revised 6 December 2018; accepted 18 January 2019; first published online 20 March 2019)

Abstract

To date, cyber security research is built on observational studies involving macro-level attributes as causal factors that account for state behaviour in cyberspace. While this tradition resulted in significant findings, it abstracts the importance of individual decision-makers. Specifically, these studies have yet to provide an account as to why states fail to integrate available information resulting in suboptimal judgements such as the misattribution of cyber operations. Using a series of vignette experiments, the study demonstrates that cognitive heuristics and motivated reasoning play a crucial role in the formation of judgements *vis-à-vis* cyberspace. While this phenomenon is frequently studied relative to the physical domain, it remains relatively unexplored in the context of cyberspace. Consequently, this study extends the existing literature by highlighting the importance of micro-level attributes in interstate cyber interactions.

Keywords: Cyber Security; Conflict; Attribution; Experiments; Bias

Introduction

Over the past decade, degradative cyber operations coincided with ongoing or emergent interstate disputes. From the Estonian and Russian debacle over a Second World War memorial, to Stuxnet and the Iranian nuclear programme, and more recently the disruption of the Ukrainian power grid, its strategic utility rests on its potential to threaten an adversary's cyber-enabled objectives.¹ Advancements in cyber capabilities, however, do not translate into strategic gains. Less than 5 per cent of observed cyber operations influenced interstate relations in favour of the aggressor.² Furthermore, rather than a surge in debilitating cyber operations, previous incidents occurred below the threshold for war and appear to demonstrate restraint³ on the part of the aggressors.⁴ One is then left to question the confidence in the domain's revolutionary potential relative to other instruments of power.⁵

Despite the above, there is no shortage of alarm following the disclosure of new cyber operations. Those framed as existential threats are often attributed to state or state-associated actors

¹Erica D. Borghard and Shawn W. Loneragan, 'The logic of coercion in cyberspace', *Security Studies*, 26:3 (2017), pp. 452–81.

²Brandon Valeriano and Ryan Maness, 'The dynamics of cyber conflict between rival antagonists, 2001–11', *Journal of Peace Research*, 51:3 (2014), pp. 347–60.

³The Stability-Instability Paradox is thought to occur with respect to the use of cyber operations by state actors.

⁴Adam P. Liff, 'Cyberwar: a new absolute weapon? The proliferation of cyberwarfare capabilities and interstate war', *Journal of Strategic Studies*, 35:3 (2012), pp. 422–6; Jon Lindsay and Erik Gartzke, 'Coercion through cyberspace: the stability-instability paradox revisited', in Kelly Greenhill and Peter Krause (eds), *The Power to Hurt: Coercion in the Modern World* (New York: Oxford University Press, 2018), p. 184; Brandon Valeriano and Ryan Maness, *Cyber War Versus Cyber Realities: Cyber Conflict in the International System* (New York: Oxford University Press, 2015), pp. 45–77.

⁵There are increasing arguments that call for its use in conjunction with other foreign policy instruments. This, however, is not in scope for this study.

and involve salient disputes between rivals.⁶ While these interactions have yet to escalate into the physical domain, states are quick to implicate rivals despite the absence of definitive evidence.⁷ This behaviour is puzzling given the rarity of cyber operations and the material, technological, and organisational resources required.⁸ Consequently, we are prompted to ask why do states believe that cyber operations are more common than they are? Furthermore, why is there a belief that rival states are more likely to resort to cyber operations as means to meet strategic ends?

In response, this study argues that this apparent deviation from rationality stems from the decision-maker's failure to update existing beliefs with available information concerning the culpability of a suspected aggressor. It posits that this non-Bayesian approach to judgement is a function of pre-existing beliefs shaped by established rivalry interactions, the responsibilities assigned to decision-makers, and the cognitive load required to process available information. To support these propositions the study employs two between-subject Internet-based vignette experiments⁹ to surface the occurrence of suboptimal judgements involving the attribution of cyber operations. These emphasise micro-level factors that are often overlooked with respect to interstate interactions in cyberspace.¹⁰

With these in mind the remainder of the article is organised as follows. The succeeding section presents the underlying theoretical framework that informs the study. This is followed by the design and methodology, which highlights the structure and limitations of the experiments. Following this, the reader is presented with the results and is provided with the general discussion of the findings. Finally, the implications of the study are surfaced with the direction that future research may take.

Before proceeding further, it is important to note that the study does not attempt to argue the novelty of micro-level factors. It, instead, demonstrates the applicability of existing frameworks in attempts to study events within cyberspace. In doing so it challenges the pervasive narrative of cyber exceptionalism to move the current discourse forward towards an empirically and theoretically grounded direction.

Theoretical framework

Non-Bayesian judgements

Alarm over cyber operations are inflated by the possible consequences resulting from the malicious disruption of critical infrastructure. While the term encapsulates a host of actions that vary in scope and severity, only a subset of these are viewed as legitimate national security concerns. While a universally accepted taxonomy of cyber operations remains elusive, incidents are classified¹¹ as either espionage, disruptive, or degradative cyber operations.¹²

Whereas espionage in the conventional sense is well understood and needs no further elaboration; cyber espionage takes advantage of the increased utilisation of the domain in support of strategic goals. Cyber espionage involves the exploitation of flaws in the underlying infrastructure

⁶Ryan Maness and Brandon Valeriano, 'The impact of cyber conflict on international interactions', *Armed Forces & Society*, 42:2 (2016), p. 305.

⁷Mischa Hansel, 'Cyber-attacks and psychological IR perspectives: Explaining misperceptions and escalation risks', *Journal of International Relations and Development*, 21:4 (2016), p. 534.

⁸Borghard and Lonergan, 'The logic of coercion in cyberspace'; Rebecca Slayton, 'What is the cyber offense-defense balance? Conceptions, causes, and assessment', *International Security*, 41:3 (2017), pp. 72–109.

⁹The experiment was registered via EGAP prior to execution and analysis. Details are available at: <http://egap.org/content/sound-alarm-bias-and-consequences-cyber-risk>.

¹⁰Rose McDermott, Jonathan Cowden, and Cheryl Koopman, 'Framing, uncertainty, and hostile communications in a crisis experiment', *International Society of Political Psychology*, 23:1 (2002), pp. 50–3.

¹¹It is useful to note though that some of these may overlap with one another to achieve a desired tactical or strategic objective.

¹²Brandon Valeriano, Benjamin Jensen, and Ryan Maness, *Cyber Strategy: The Evolving Character of Power and Coercion* (New York: Oxford University Press, 2018), pp. 22–52.

to exfiltrate privileged information. Moreover, these enable later, and more aggressive, operations.¹³ Disruptive cyber operations, in contrast, typically manifest themselves through defacement and (distributed) denial-of-service (DoS) against an adversary. These require fewer resources than the former and are perceived as a nuisance given its transient and limited effects. Most interstate exchanges in cyberspace are categorised as either espionage or disruptive cyber operations.¹⁴

Degradative cyber operations¹⁵ are characterised by technical sophistication, the organisational maturity of aggressors, and latent strategic potential. These exploit vulnerabilities that are typically out of reach for most actors (state and non-state) without the necessary competencies. While early advocates argued that cyberspace reduces material imbalances between actors, previous incidents prove otherwise. Operational outcomes depend not only on the availability of technical resources, but on the support offered by other policy instruments and enabled by mature organisational processes.¹⁶ Ultimately, the ability to threaten infrastructure crucial to an adversary's political, economic, and/or military objectives facilitates its perceived utility.

With respect to real-world policy outcomes, the significance assigned to degradative cyber operations is reflected in several national and regional strategies and initiatives. With respect to the former, the recently published United States National Cyber Strategy associates the protection of critical infrastructure with 'Protecting the American people, the American way of life...';¹⁷ effectively echoing previous rhetoric surrounding the potential impact of these types of cyber operations.¹⁸ Similarly, the Singaporean Cyber Strategy also acknowledges that operations that affect critical infrastructure may result in 'disruptions which could cripple economies, and lead to loss of life'.¹⁹ At a regional level, organisations such as ASEAN recognise the economic and societal implications of these operations.²⁰ And while it is important to note that these concerns are not necessarily shared uniformly by all political decision-makers given the unequal dependence on these technologies, continued integration of ICT in day-to-day life at a global level suggests that global perceptions could gravitate towards this shared perspective.²¹

The consequences of these operations relative to the growth of cyberspace is made starker by the domain's interconnected and interdependent nature.²² As with similarly complex constructs, the deeper and more entangled these become, the greater the difficulty to identify and rectify flaws.²³ This sense of unknowability and vulnerability is made pressing by the possible discovery

¹³Ben Buchanan, *The Cybersecurity Dilemma: Hacking, Trust and Fear Between Nations* (London: Hurst & Company, 2017), pp. 5–6.

¹⁴Council on Foreign Relations, 'Cyber Operations Tracker', available at: {<https://www.cfr.org/interactive/cyber-operations>} accessed 28 November 2018; Valeriano and Maness, 'The dynamics of cyber conflict between rival antagonists, 2001–11'.

¹⁵Moving forward, all references to 'cyber operations' or 'operations' refer to 'degradative cyber operations'.

¹⁶Borghard and Lonergan, 'The logic of coercion in cyberspace', p. 474; Lindsay and Gartzke, 'Coercion through cyberspace', p. 173; Slayton, 'What is the cyber offense-defense balance?', pp. 87–91.

¹⁷United States of America, 'National Cyber Security Strategy of the United States of America', available at: {<https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf>} accessed 28 November 2018.

¹⁸Elisabeth Bumiller and Thom Shanker, 'Panetta warns of dire threat of cyberattack on U.S.', *The New York Times*, available at: {<https://www.nytimes.com/2012/10/12/world/panetta-warns-of-dire-threat-of-cyberattack.html>} accessed 28 November 2018.

¹⁹Cyber Security Agency of Singapore, 'Singapore Cybersecurity Strategy', available at: {[https://www.csa.gov.sg/~media/csa/documents/publications/singaporecybersecuritystrategy.pdf](https://www.csa.gov.sg/~/media/csa/documents/publications/singaporecybersecuritystrategy.pdf)} accessed 28 November 2018.

²⁰Elena L. Aben, 'ASEAN to form cybersecurity group', *Manila Bulletin*, available at: {<http://www.pressreader.com/philippines/manila-bulletin/20160528/281547995138115>} accessed 28 November 2018.

²¹Miguel Alberto Gomez and Candice Tran Dai, 'Challenges and opportunities for cyber norms in ASEAN', *Journal of Cyber Policy*, 3:2 (2018), pp. 217–35.

²²Martin Libicki, 'Cyberdeterrence and Cyberwar', RAND Corporation, available at: {https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf} accessed 29 November 2018.

²³Charles Perrow, *Normal Accidents: Living with High Risk Technologies* (New Jersey: Princeton University Press, 1984), pp. 62–100.

and exploitation by malicious actors.²⁴ Furthermore, the interconnection between components implies the potential for cascading damage across the domain.²⁵ Yet despite such, the above scenario remains unsubstantiated by past interstate cyber operations.

None of the degradative cyber operations observed to date are thought to have been 'strategically successful'. While there is no doubt that some may be classified as tactically promising, none have managed to influence the behaviour of their respective targets in the manner intended by aggressors.²⁶ And while Brandon Valeriano, Benjamin Jensen, and Ryan Maness²⁷ identify a handful of successful operations, it is difficult to ascertain whether these were the result of independent use or joint employment with other foreign policy instruments.²⁸

Few states possess the technical capabilities required for these operations.²⁹ While cyberspace facilitates the diffusion of power, it does not follow that material imbalances are made irrelevant.³⁰ A degree of organisational maturity is required to effectively utilise latent cyber power.

Although the proliferation of cyber capabilities enables actors to operate within this domain, their ability to marshal their resources is just as important. Rebecca Slayton³¹ notes that the strategic potential of cyberspace is only realised through the careful use of skills, intelligence, and technology. The outcome of cyber operations is therefore determined by the ability to access and integrate these resources effectively.³²

Beyond technological and organisational considerations, the use of degradative cyber operations are limited by the need for restraint. Valeriano and Maness³³ observe that even with these requirements in their possession, states willingly engage in less effective operations. Citing the need to avoid crossing 'red lines', the authors argue that restraint serves to minimise the risk of escalation. And while there has yet to be a case wherein violence in cyberspace has carried over to the physical domain,³⁴ Ben Buchanan³⁵ notes that the uncertainty surrounding the intent of these operations encourages such concerns. In a hypothetical attack against a state's nuclear command, control, and communications (NC3), Erik Gartzke and Jon Lindsay³⁶ foresee an escalation spiral. Complementing this finding, Jacquelyn Schneider,³⁷ in a series of war games, demonstrates that the possibility of escalation is indeed recognised by elite decision-makers when faced with the option of utilising cyber operations.

²⁴Myriam Dunn Cavelty, 'From cyber-bombs to political fallout: Threat representations with an impact in the cyber-security discourse', *International Studies Review*, 15:1 (2013), pp. 114–15.

²⁵Ilai Saltzman, 'Cyber posturing and the offense-defense balance', *Contemporary Security Policy*, 34:1 (2013), pp. 40–63.

²⁶Jason Healey, 'Winning and losing in cyberspace', in Nikolaos Pissanidis, Henry Røigas, and Matthijs Veenendaal (eds), *8th International Conference on Cyber Conflict* (Tallinn: IEEE, 2016), pp. 37–49; Emilio Iasiello, 'Cyber attack: a dull tool to shape foreign policy', in Karlis Podins, Jan Stinissen, and Markus Maybaum (eds), *5th International Conference on Cyber Conflict* (Tallinn: IEEE, 2013), pp. 451–70; Maness and Valeriano, 'The impact of cyber conflict on international interactions', pp. 301–23.

²⁷Valeriano, Jensen, and Maness, *Cyber Strategy*, pp. 22–52.

²⁸Furthermore, the extent with which cyber operations contributed to its success is also in doubt. This reflects the growing trend of considering cyber operations as one component of a larger strategic campaign.

²⁹Allison Pytlak and George Mitchell, 'Power, rivalry, and cyber conflict: an empirical analysis', in Karsten Friis and Jens Ringsmore (eds), *Conflict in Cyber Space: Theoretical, Strategic and Legal Perspectives* (London: Routledge, 2016), pp. 65–82; Slayton, 'What is the cyber offense-defense balance?', pp. 72–109.

³⁰Joseph Nye, 'Cyber Power', Belfer Center for Science and International Affairs, available at: {<https://www.belfercenter.org/sites/default/files/legacy/files/cyber-power.pdf>} accessed 29 November 2018.

³¹Slayton, 'What is the cyber offense-defense balance?', pp. 72–109.

³²Liff, 'Cyberwar', pp. 409–21.

³³Valeriano and Maness, *Cyber War Versus Cyber Realities*, pp. 45–77.

³⁴Nadiya Kostyuk and Yuri Zhukov, 'Invisible digital front: Can cyber attacks shape battlefield events?', *Journal of Conflict Resolution* (2017).

³⁵Buchanan, *The Cybersecurity Dilemma*, p. 20.

³⁶Erik Gartzke and Jon Lindsay, 'Thermonuclear cyberwar', *Journal of Cybersecurity*, 3:1 (2017), p. 45.

³⁷Jacquelyn Schneider, 'Cyber and Crisis Escalation: Insights from Wargaming', US Naval War College, available at: {<https://pacs.einaudi.cornell.edu/sites/pacs/files/Schneider.Cyber%20and%20Crisis%20Escalation%20Insights%20from%20Wargaming%20Schneider%20for%20Cornell.10-12-17.pdf>} accessed 29 November 2018.

In lieu of the technological, organisational, and strategic constraints that limit the exercise of degradative cyber operation, why are these perceived to be a frequent occurrence by policymakers and popular media? Authors such as Richard Clarke and Robert Knake³⁸ regard it as a real and imminent threat to national security. Relatedly, a study of news articles from 2008 to 2013 highlights the elevated levels of concern in response to increasingly complex cyber operations.³⁹ Similarly, a recent survey in the United States indicates that 73 per cent of Americans perceive 'cyber terrorism' as a 'critical threat' to the United States.⁴⁰ Finally, the number of states instituting national cyber strategies has more than doubled since 2007⁴¹ with interest in the politico-strategic implications of cyber conflict emerging in response to notable incidents in cyberspace.⁴²

Much like terrorism and other High-Impact Low-Probability Events (HILP), normative strategies in assessing risk fail and lead to an increase in dread risk.⁴³ While a host of factors aggravate this deviation from rationality, its causal process is depicted simply as: Fear + Uncertainty + Media Exposure – Experience = inflated assessments.⁴⁴ When these factors are present, decision-makers are prone to resorting to cognitive processes that result in suboptimal judgements.⁴⁵ Consequently, when faced with degradative cyber operations that exhibit the above characteristics, decision-makers are prone to disregarding probabilities through a combination of motivated reasoning and cognitive limitations; resulting in suboptimal judgements. As such:

Hypothesis 1. The misuse of probabilities results in misattribution when responding to degradative cyber operations

Cyber operations often involve states mired in rivalry over salient issues. On the one hand, these relationships frame these interactions as extensions of pre-existing rivalry behaviour and are, thus, stable.⁴⁶ On the other hand, the sudden appearance of novel offence-oriented capabilities may trigger a security dilemma that alters the rivalry dynamics between states.⁴⁷ This possible shift is worrisome as most states with the capacity to invest in cyber operations are both economically and militarily capable. At present, only the United States, China, Iran, Russia, and North Korea appear to demonstrate proficiency in these types of operations.⁴⁸ But given that these are notable conventional (and nuclear) powers, an unforeseen escalation is worrisome. Misattribution stemming from the incorrect use (or dismissal) of probabilities may, at best, result in the development

³⁸Richard Clarke and Robert Knake, *Cyber War* (New York: Harper-Collins, 2010), pp. 30–1.

³⁹Lee Jarvis, Stuart MacDonald, and Andrew Whiting, 'Unpacking cyberterrorism discourse: Specificity, status, and scale in news media constructions of threat', *European Journal of International Security*, 2:1 (2017), pp. 64–87.

⁴⁰Justin McCarthy, 'Americans cite cyberterrorism among top three threats to U.S.', *Gallup*, available at: {<https://news.gallup.com/poll/189161/americans-cite-cyberterrorism-among-top-three-threats.aspx>} accessed 29 November 2018.

⁴¹NATO Cyber Defence Centre of Excellence, 'Cyber Security Strategy Documents', available at: {<https://ccdcoc.org/cyber-security-strategy-documents.html>} accessed 29 November 2018.

⁴²This is not to say, however, that a uniform perception of threat exists amongst these states.

⁴³Belief in the occurrence of a low-probability, high-damaging event that occurs at a given point in time; Colin Camerer and Howard Kunreuther, 'Decision-processes for low probability events – policy implications', *Journal of Policy Analysis and Management*, 8:4 (1989), p. 565.

⁴⁴Gina Reinhardt, 'Imagining worse than reality: Comparing beliefs and intentions between disaster evacuees and survey respondents', *Journal of Risk Research*, 20:2 (2017), pp. 169–94.

⁴⁵Camerer and Kunreuther, 'Decision-processes for low probability events', pp. 565–92; Gerd Gigerenzer, 'Out of the fry-ing pan into the fire: Behavioral reactions to terrorist attacks', *Risk Analysis*, 26:2 (2006), pp. 347–51; Kip Viscusi and Richard Zeckhauser, 'Recollection bias and its underpinnings: Lessons from terrorism risk assessments', *Risk Analysis*, 37:5 (2017), pp. 969–81.

⁴⁶Valeriano and Maness, *Cyber War Versus Cyber Realities*, pp. 45–77.

⁴⁷Buchanan, *The Cybersecurity Dilemma*, pp. 75–100; Gartzke and Lindsay, 'Thermonuclear cyberwar', pp. 44–5.

⁴⁸Valeriano and Maness, 'The dynamics of cyber conflict between rival antagonists, 2001–11', pp. 347–60.

of inappropriate policies while, at worst, lead to an escalatory spiral. The latter is potentially escalatory in cases where targets are assets crucial to a state's strategic or ideological interests.⁴⁹

Since attribution is ultimately a cognitive process where both technical and contextual evidence informs judgement, its outcome is best understood at this level of analysis. Broadly speaking, the study posits that an interaction between hot and cold cognitive processes results in the misuse of probabilities. Hot cognition is characterised by judgements that are affect-driven and manifests as the need to maintain existing beliefs. In contrast, cold cognition occurs independent of affect and is associated with individual cognitive limitations.⁵⁰

The motivated thinker

The notion of a rational actor requires the identification of a solution only once all possible alternatives and outcomes have been evaluated. In this idealised scenario, supporting and contradictory information are used to update existing beliefs. Since Jervis's seminal work, scholars have argued that the availability of new information serves to either reaffirm pre-existing beliefs if congruent or are ignored if contradictory.⁵¹

Beliefs originate from the need to understand the environment and enables certain cognitive processes.⁵² One of which is to set or influence expectations. For instance, if you routinely observe your neighbour's dog covered in mud and leaving muddy paw prints on the side walk, you associate the presence of these paw prints with the dog. If it occurs frequently, the idea that your neighbour's dog enjoys walking around in mud reinforces this belief. In the future, when the dog is not seen but the paw prints are present, this belief allows one to accept the plausibility of certain propositions (for example, the muddy paw prints came from the same muddy dog) such that judgements that emerge reinforce this existing belief (that is, the same dog made those prints) at the cost of alternative explanations (another dog may have made those prints). Adapting this argument to interstate relations, events such as the Yom Kippur War are indicative of pre-existing beliefs setting expectations.⁵³

Researchers posit that motivated reasoning is rooted in past affect-laden experiences. When a decision-maker processes new information, they depend on affect-laden information stored in long-term memory. Once accessed, a specific affect is triggered along with the associated information. A heuristic mechanism evaluates these feelings with respect to the information triggered and reinforces the existing affect irrespective of dis-confirmatory information.⁵⁴ Jonathan

⁴⁹Chris Whyte, 'Ending cyber coercion: Computer network attack, exploitation and the case of North Korea', *Comparative Strategy*, 35:2 (2016), pp. 93–102.

⁵⁰Dennis Chong, 'Degree of rationality in politics', in Leonie Huddy and David Sears (eds), *The Oxford Handbook of Political Psychology* (New York: Oxford University Press, 2013), pp. 96–129.

⁵¹Richard Herrmann, James Voss, Tonya Schooler, and Joseph Ciarrochi, 'Images in International Relations: an experimental test of cognitive schemata', *International Studies Quarterly*, 41:3 (1997), pp. 402–33; Marcus Holmes, 'Believing this and alieving that: Theorizing affect and intuitions in international politics', *International Studies Quarterly*, 59:4 (2015), pp. 706–20; Robert Jervis, *Perception and Misperception in International Politics* (New Jersey: Princeton University Press, 1976), pp. 13–31; Robert Jervis, 'Understanding beliefs and threat inflation', in Trevor Thrall and Jane Cramer (eds), *American Foreign Policy and the Politics of Fear* (New York: Routledge, 2009), pp. 16–39; Jonathan Mercer, 'Emotional beliefs', *International Organization*, 64:1 (2010), pp. 1–31; Steven Roach, 'Affective values in international relations: Theorizing emotional actions and the value of resilience', *Politics*, 36:4 (2016), pp. 400–12; Brent Sasley, 'Affective attachments and foreign policy: Israel and the 1993 Oslo Accords', *European Journal of International Relations*, 16:4 (2010), pp. 687–709.

⁵²Jervis, *Perception and Misperception in International Politics*, pp. 13–31.

⁵³Uri Bar-Joseph and Arie Kruglanski, 'Intelligence failure and need for cognitive closure: On the psychology of the Yom Kippur surprise', *Political Psychology*, 24:1 (2003), pp. 75–99.

⁵⁴Milton Lodge and Charles Taber, 'Three steps towards a theory of motivated political reasoning', in Arthur Lupia, Matthew McCubbins, and Samuel Popkin (eds), *Elements of Reason: Cognition, Choice, and the Bounds of Rationality* (New York: Cambridge University Press, 2000), pp. 183–213; Charles Taber, Milton Lodge, and Jill Glathar, 'The motivate

Mercer⁵⁵ extends this argument and posits that emotions function as an assimilative mechanism for beliefs. Whereas rationalists assume that new information serves to update existing beliefs and eventually converges towards reality, Mercer argues that emotions assimilate data into beliefs; thus reinforcing it. These echo Jervis's rationale for maintaining beliefs and provides the necessary psychological basis for this argument. These beliefs are, in turn, reflected through images held by decision-makers.

Images are the 'total cognitive, affective, and evaluative structure of the behavior unit or its internal view of itself and its universe'.⁵⁶ This links the construct of images with both affect and belief. With respect to state behaviour, images serve three different tasks: (1) the evaluation of the relative capability of another actor; (2) the perceived threat and/or opportunity given to another actor; and (3) the perceived culture of another actor. The study focuses on the second function and the decision-maker's need to maintain an 'enemy image' of a threatening other.⁵⁷

Enemy images are constructs through which other actors are perceived to behave in 'bad faith'.⁵⁸ While several factors result in the emergence of enemy images, the study focuses on conflict accumulation as a reinforcement mechanism. David Dreyer⁵⁹ notes that constant exposure to multiple issues in a rivalry environment reinforces these images and increases the likelihood conflict. Crucial to this notion is the idea that an enemy image is strengthened irrespective of subsequent issues (that is, the dispute need not be over related actions or issues). This lack of differentiation is unsurprising as individuals only recall basic concepts of past events known as gist.⁶⁰ This absence of context allows repeated exposure to negative interactions to strengthen pre-existing images.⁶¹ Consequently, if the notion that state behaviour in cyberspace is a reflection of rivalry dynamics, it is likely that an accumulative process that strengthens these images is at work.⁶² More worrying, the need to maintain an existing enemy image may further be aggravated by existing processes that may result in attribution in support of a political interest among a small group of decision-makers. As noted by Thomas Rid and Ben Buchanan,⁶³ the attribution of cyber operations includes a strategic component during which decision-makers take into consideration their understanding of an adversary's motivation and past behaviour. Given the need to maintain pre-existing beliefs, it follows that:

Hypothesis 2.1. Decision-makers are likely to ignore probabilities to support information aligned with pre-existing enemy images

The need to maintain beliefs cannot exclusively account for suboptimal judgements. The need for cognitive closure driven by associated costs among elites is equally important.⁶⁴ For elites, associated

construction of political judgements', in James Kuklinski (ed.), *Citizens and Politics: Perspectives from Political Psychology* (New York: Cambridge University Press, 2001), pp. 198–226.

⁵⁵Mercer, 'Emotional beliefs', p. 9.

⁵⁶Kenneth Boulding, 'National images and international systems', *Journal of Conflict Resolution*, 3:2 (1959), pp. 121–2.

⁵⁷Both capabilities and cultural likeness may be significant to state interactions in cyberspace but are beyond the scope of this study.

⁵⁸Ole Holsti, 'The belief system and national images: a case study', *Journal of Conflict Resolution*, 6:3 (1962), p. 247; Ole Holsti, 'Cognitive dynamics and images of the enemy', *Journal of International Affairs*, 21:1 (1967), p. 17.

⁵⁹David Dreyer, 'Issue conflict accumulation and the dynamics of strategic rivalry', *International Studies Quarterly*, 54:3 (2010), pp. 779–95.

⁶⁰Only a general idea of a previous event is retained in memory, which could result in mis-contextualisation when used in the future.

⁶¹Scott Blum, Roxane Silver, and Michael Poulin, 'Perceiving risk in a dangerous world: Associations between life experiences and risk perceptions', *Social Cognition*, 32:3 (2014), pp. 299–300; Dreyer, 'Issue conflict accumulation', pp. 784–6.

⁶²Maness and Valeriano, 'The impact of cyber conflict on international interactions', p. 305.

⁶³Thomas Rid and Ben Buchanan, 'Attributing cyber attacks', *Journal of Strategic Studies*, 38:1 (2015), p. 25.

⁶⁴Arie Kruglanski and Donna Webster, 'Motivated closing of the mind: Seizing and freezing', *Psychological Review*, 103:2 (1996), pp. 68–111.

costs are the function of a decision-maker's role and accountability.⁶⁵ This is particularly true in democratic regimes where elites are held accountable for policy failures. Self-interest motivates elites to maintain existing beliefs at the expense of possible alternatives.⁶⁶ Consequently, closure must occur as quickly as possible (urgency) and must last for as long as possible⁶⁷ (permanence).

Urgency encourages decision-makers to 'seize' upon cues that provide immediate closure. For instance, events that appear as a rival's attempt to obtain an advantage are believed to be so. Once seized, decision-makers 'freeze'⁶⁸ on their judgements when presented with contradictory information. Building on the preceding example, even if a rival's actions are later deemed to be benign, this information is disregarded in favour of maintaining an existing belief. As such:

Hypothesis 2.2. Decision-makers in positions of authority are likely to utilise information that results in immediate closure

The miserly processor

Other non-affective processes are equally informative in determining the tendency of decision-makers to misuse probabilities. Although the underlying political environment suggests a greater role for motivated cognition, the inherent characteristics of cyberspace impose substantial cognitive constraints.⁶⁹ Lene Hansen and Helen Nissenbaum⁷⁰ assert that the knowledge-gap between technology experts and policymakers resulted in threat inflation among the latter. Additionally, Slayton⁷¹ notes that a lack of technical understanding regarding the difficulties associated with developing offensive campaigns in cyberspace fuels the myth of an offensive advantage within the domain. These biased judgements are less a function of motivated thinking and more a result of inherent cognitive limitations. For this study, suboptimal judgements are believed to occur in part due to the task-difficulty and limited cognitive capacity that constrains processing motivation.⁷² Processing motivation does not refer to the maintenance of specific beliefs nor does it denote efforts to reach a predefined conclusion. It pertains instead to the amount of cognitive effort required to accomplish a task.

The processes and structures of cyberspace are highly abstracted for the benefit of non-experts. On the one hand, this permits the utilisation and expansion of the domain without the need to comprehend its inner workings. On the other, an objective analysis of security incidents requires a nuanced understanding of cyberspace. Together with the politico-strategic milieu of interstate cyber interactions, these burden the cognitive capacities of decision-makers. In response, cognitive shortcuts are invoked to formulate judgements using the least amount of (cognitive) resources. And while cognitive heuristics have been proven to assist experts in formulating judgements, these may promote biases and suboptimal judgements.⁷³ While several factors⁷⁴ enable these biases, the study focuses on information order.

⁶⁵Jennifer Lerner and Philip Tetlock, 'Accounting for the effects of accountability', *Psychological Bulletin*, 125:2 (1999), pp. 255–75.

⁶⁶Chong, 'Degree of rationality in politics', pp. 96–129.

⁶⁷Other aspects such as personality and leadership style may also severely impact the extent to which information is processed but these are not currently in scope; Bar-Joseph and Kruglanski, 'Intelligence failure and need for cognitive closure', p. 81.

⁶⁸While significant to the study of bias, the freezing effect is not tested explicitly in these experiments.

⁶⁹Rid and Buchanan, 'Attributing cyber attacks', p. 5.

⁷⁰Lene Hansen and Helen Nissenbaum, 'Digital disaster, cyber security, and the Copenhagen School', *International Studies Quarterly*, 53:4 (2009), pp. 1155–75.

⁷¹Slayton, 'What is the cyber offense-defense balance?', pp. 72–109.

⁷²Bar-Joseph and Kruglanski, 'Intelligence failure and need for cognitive closure', pp. 75–99.

⁷³Richard Lau and David Redlawsk, 'Advantages and disadvantages of cognitive heuristics in political decision making', *American Journal of Political Science*, 45:4 (2001), pp. 951–71.

⁷⁴Factors such as task-specific instructions, length, and response format have been shown to influence the extent to which base rates are ignored.

There is no means to guarantee the order in which relevant information is obtained in response to malicious activities in cyberspace. While frameworks exist to improve this process, analysts depend on crucial evidence wherever and whenever available. And while it would be ideal to treat each piece of evidence objectively, recently acquired information may have greater saliency due to the limited capacity of working memory. The demonstration of this *recency effect* confirms the tendency of individuals to prefer the path of least cognitive resistance.⁷⁵ With respect to probabilities, experiments illustrate the tendency to shy away from cognitively strenuous computation of actual probabilities in favour of using the most recent information provided. As such:

Hypothesis 3. Decision-makers are influenced by the order of information and are likely to privilege the most recent information received

While the importance of motivated reasoning and cognitive heuristics in International Relations and political science has been studied for quite some time, there is no reason to believe that the novelty of cyberspace nullifies lessons learned. While there is no doubt that the nature of cyberspace is distinct from that of the physical domain, there is no reason to believe that existing theories and frameworks are inapplicable. Given the centrality of human actors, it is more than likely that cognitive processes provide crucial insight into state behaviour in cyberspace.

Research design and methodology

Operationalisation

The study utilises two between-subject Internet-based vignette survey experiments to test the proposed framework. The vignette depicts an ongoing territorial dispute between two states (Country A and Country B) involving resource-rich islands in contested waters.⁷⁶ In the vignette, Country A's claim is recognised by the international community. At a certain point in time, Country A's offshore rigs are subjected to a cyber operation that disables these facilities for several hours. After which participants are exposed to treatments corresponding to the relevant variables⁷⁷ in this study.⁷⁸

As noted earlier, motivated reasoning emerges as a function of enemy images (*Image*) and existing organisational role (*Role*). The treatment for images is applied by manipulating Country B's past behaviour. An enemy image is reflected in past non-cooperative and martial behaviour. Correspondingly, a non-enemy image is present when efforts to reach a compromise are observed. The latter serves as the control for this treatment. To test for the impact of organisational role, participants are assigned one of two roles. The first is that of the administrator of Country A's National Cyber Command (NCC) responsible for advising the head of state as to the appropriate policy response. The second is that of one of many analysts responsible for providing his or her superiors with the necessary information. The latter serves as the control for this treatment. *Role* is interacted with both *Image* and *Order* to test for the effect of seizing. Participants playing the role of the NCC administrator are more likely to accept confirmatory evidence.

In testing for non-motivated reasoning, the study manipulates the order in which the information is presented (*Order*). The original ordering presents the base rate information before individuating information. This serves as the control for the treatment. To test for the impact of information order the sequence is reversed. If decision-makers are indeed susceptible to the

⁷⁵Jon Krosnick, Fan Li, and Darrin Lehman, 'Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments', *Journal of Personality and Social Psychology*, 59:6 (1990), pp. 1140–52.

⁷⁶The scenario is loosely based on the ongoing dispute between China and several Southeast Asian states.

⁷⁷When terms are italicised, these refer to specific variables being studied.

⁷⁸See Appendix.

sequence associated with information acquisition, later information is prioritised compared to those presented earlier.

After participants are presented with these details, they provide an assessment as to whether Country B is responsible for the cyber operation. Their response is collected in the form of a probability (*Probability*) from 0–100 per cent.⁷⁹

External and internal validity issues

As with other experimental studies, issues of external and internal validity are often a point of contention. Chief among these are concerns over the study's external validity. To begin with, the study does not claim generalisability. These results are only applicable with the sample in hand and generalisation beyond this is discouraged. Moreover, with experiments involving elite decision-making, claims made must be taken with a degree of caution given the underlying differences between elites and non-elite samples.⁸⁰

To mitigate issues concerning external validity the study employs two separate samples: students and government employees.⁸¹ This considers the possible variation between these two groups. In addition, the study utilises the Internet to recruit participants to provide a wider demographic reach that serves to mitigate any culture-specific effects.⁸² Finally, the design is strengthened by providing a scenario that integrates aspects of real-world events in an attempt to increase the impact of this fictitious scenario.

Besides external validity, questions of internal validity need to be addressed as well. Two stand out in this regard: the pre-treatment effect and engagement with the study. The former refers to the possibility that participants may already possess knowledge related to the study that may skew the results towards their pre-existing beliefs. This may come about due to exposure (directly or indirectly) to similar cases or through formal education or first-hand experience. While limiting contact with similar events is unlikely given the media coverage surrounding cyber operations, it is possible to address this by not restricting participation to individuals from a specific educational background thus increasing variation within the samples.

The issue of engagement, in comparison, requires a more stringent approach. Participants of Internet-based experiments tend to be disengaged with the material (that is, less attentive). This limits the effectiveness of the intended treatment. To overcome this issue, four (4) attention check questions are used within the experiment. Failure to correctly answer all four questions results in the exclusion of a participant. Previous research demonstrates that Internet-based experiments can expect an exclusion rate of 30–50 per cent.⁸³

Participant recruitment

The study recruits its participants through the online platform *Prolific*. While the debate continues regarding the validity of results obtained from Internet-based experiments, research shows no

⁷⁹Although it has been shown that the response format can influence the emergence of bias, this is not the primary concern of this study and is consequently not tested.

⁸⁰Alex Mintz, Steven Redd, and Arnold Vedlitz, 'Can we generalize from student experiments to the real world in political science, military affairs, and international relations?', *Journal of Conflict Resolution*, 50:5 (2006), pp. 757–76.

⁸¹In light of the difficulty of acquiring political elites for these experiments, this serves as a viable proxy for the purpose of this study.

⁸²John Aldrich and Arthur Lupia, 'Experiments and game theory's value to political science', in James Druckman, Donald Green, James Kuklinski, and Arthur Lupia (eds), *Cambridge Handbook of Experimental Political Science* (New York: Cambridge University Press, 2011), pp. 89–101.

⁸³Krista Casler, Lydia Bickel, and Elizabeth Hackett, 'Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing', *Computers in Human Behavior*, 29:6 (2013), pp. 2156–60; Miguel Alberto Gomez and Eula Bianca Villar, 'Fear, uncertainty, and dread: Cognitive heuristics and cyber threats', *Politics and Governance*, 6:2 (2018), pp. 61–72.

significant difference between this and conventional lab-based environments.⁸⁴ However, issues pertaining to external and internal validity demands care during participant recruitment.

To begin with, the study is conducted using two separate groups. The first consisting of undergraduate students and the other of government employees.⁸⁵ The former represents the usual approach in these types of experiments.⁸⁶ The latter, however, is used to allay fears concerning the non-representativeness of the sample.⁸⁷ Given the difficulty of recruiting elites, government employees serve as a readily accessible proxy.

For both groups, additional steps are taken to ensure the rigour of the design. First, only those with English as their primary language are invited to participate. This addresses concerns regarding comprehension that may alter the accuracy of the response. Second, participants are not selected based on a specific educational background. This increases the variation between those with greater knowledge about cyberspace and those with less. Finally, only those with an approval rating of 90 per cent or higher are accepted. *Prolific* grades possible participants according to the number of times their responses have been accepted by other researchers. This aids in reducing the engagement problem discussed earlier.

Once the participants are selected, they are randomly assigned to one of eight (8) possible scenarios that correspond to the manipulation of *Image*, *Role*, and *Order*. Upon completing the experiment, participants are rewarded US \$0.35 for their participation in the three to seven-minute-long activity.

Key limitations

Before proceeding further, the limitations of this design must be surfaced to better frame the context in which the results are interpreted. First among these is the limited generalisability of the findings. While experimental designs allow unprecedented control over the variables of interest, it comes at the cost of wider and more complex processes that readily occur in the social world.⁸⁸ Although the manipulation of *Image*, *Role*, and *Order* permits us to establish their function in the formation of judgements, this comes at the cost of other externalities. Moreover, the manipulation of these variables requires the (over) simplification of real-world processes that results in a much narrower analysis of this phenomenon.

Second, despite attempts to recruit individuals more closely aligned with those expected to face similar judgmental tasks, there is no guarantee that participants will behave in a similar fashion. Key differences are noted between elites and non-elites with regards to the formation of judgements. Elites are believed to be: (1) less prone to loss aversion; (2) may be better at multi-stage thinking; (3) able to utilise heuristics more effectively; (4) prone to overconfidence; (5) and are thought to be more cooperative.⁸⁹ Having participants that better fit the notion of non-elites would result in findings that do not account for the behaviour of the population of interest.

Third, foreign policy judgements are rarely, if ever, made in isolation. These are subject to bureaucratic and organisational processes that either enforce or temper individual idiosyncrasies.

⁸⁴Casler, Bickel, and Hackett, 'Separate but equal', pp. 2156–60; Matthew Crump, John McDonnell, and Todd Gureckis, 'Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research', *Plos One*, 8:3 (2013), p. e57410; Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti, 'Beyond the Turk: Alternative platforms for crowdsourcing behavioral research', *Journal of Experimental Social Psychology*, 70 (2017), pp. 153–63.

⁸⁵Prolific allows researchers to screen participants based on predefined criteria. There is, however, no guarantee that individuals were truthful when they provided the required background information.

⁸⁶Mintz, Redd, and Vedlitz, 'Can we generalize ...?', pp. 757–76.

⁸⁷Emilie Hafner-Burton, Alex Hughes, and David Victor, 'The cognitive revolution and the political psychology of elite decision making', *Perspectives on Politics*, 11:2 (2013), pp. 368–86.

⁸⁸Rose McDermott, *Political Psychology in International Relations* (Michigan: University of Michigan Press, 2004), pp. 24–9.

⁸⁹Hafner-Burton, Hughes, and Victor, 'The cognitive revolution', pp. 370–3.

This study, however, avoids this process in the interest of analytical parsimony. At most, the importance of the organisation is collapsed into the construct of *Role* and its influence on individual judgement. In doing so, the study is unable to elaborate on the persistence of biases once individual perspectives are subjected to organisational and bureaucratic processes.

Fourth, the study lacks the dynamic nature of real-world cases of cyber operations. Information is not static and the process of obtaining evidence may, itself, be subject to a separate set of biases that ultimately skews judgements concerning accountability. For instance, evidence may be collected by an organisation in a manner to support its own pre-held conceptions of reality. Consequently, this skews the judgement made by the information consumer. The design, as it stands, is unable to account for such.

Yet despite the above limitations, the findings surfaced by this study are not without merit. Instead, these constraints guide the reader as how best to interpret the results presented in the succeeding section. With respect to other scholars, these limitations invite further studies that could either support or refute conclusions that emerge from the present design.

Experimental results

Experiment 1. Undergraduate sample

For this experiment a total of 448 participants were recruited. Of these, 36.61 per cent (164) were removed after failing the attention check questions. To ensure a balanced analysis, random samples were drawn based on the size of the smallest treatment group resulting in 176 samples with 22 samples per treatment group. This sample had 55.11 per cent (97) male and 44.89 per cent (79) female participants with an average age of 25.46 years. Geographically, most participants were from either Europe (47.73 per cent) or the Americas (44.32 per cent). On average, participants took approximately 2 minutes and 57 seconds to complete the experiment. The average probability that Country B was responsible for the cyber operation is 21.65 per cent.

To determine the effect of *Image*, *Role*, and *Order* on *Probability*, a Type I Analysis of Variance (ANOVA) is performed. In the model, the interaction terms *Role* x *Image* and *Role* x *Order* are used to test for the presence of seizing identified in the above framework.

The results highlight a significant Average Treatment Effect (ATE) due to *Order* $F(1,163) = 4.612$ treatment on *Probability* at the $p < 0.05$ level with an effect size of $f = 0.166$. The small effect size is not exclusively associated with sample size. The extent to which extraneous factors were controlled for, the specific manipulations, and the medium used for this experiment may have reduced the magnitude, but not the significance, of the treatment's impact on participants.⁹⁰ None of the other variables and interaction terms were found to be statistically significant. A post-hoc comparison further reveals that the main effects of *Order* are significant at the $p < 0.05$ level. When individuating information is presented first, *Probability* increases by 7.58 percentage points.

Experiment 2. Government employee sample

For this experiment a total of 423 participants were recruited. Of these, 40.9 per cent (173) were removed after failing the attention check questions. To ensure a balanced analysis, random samples were drawn based on the size of the smallest treatment group resulting in 200 samples with 25 samples per treatment group. This sample had 37 per cent (74) male and 63 per cent (126) female participants with an average age of 36.49 years. Geographically, most participants were from either Europe (59.5 per cent) or the Americas (37 per cent). On average, participants took approximately 3 minutes 16 seconds to complete the experiment. The average probability that Country B was responsible for the cyber operation is 21.82 per cent.⁹¹

⁹⁰Elliot Aronson and Merrill Carlsmith, *Methods of Research in Social Psychology* (New York: McGraw-Hill, 1990), pp. 42–9.

⁹¹This is not statistically different from the first experiment.

To determine the effect of *Image*, *Role*, and *Order* on *Probability*, a Type I Analysis of Variance (ANOVA) is performed. In the model, the interaction terms *Role* x *Image* and *Role* x *Order* are used to test for the presence of seizing identified in the above framework.

The results highlight a significant Average Treatment Effect (ATE) due to *Image* $F(1,194) = 8.787$ treatment on *Probability* at the $p < 0.05$ level with an effect size of $f = 0.213$. None of the other variables and interaction terms were found to be statistically significant. A post-hoc comparison further reveals that the main effects of *Image* are significant at the $p < 0.05$ level. When an enemy image is present, *Probability* increases by 9.32 percentage points.

General discussion

While the mean of *Probability* for both experiments were not statistically different from one another ($x_1 = 21.653$, $x_2 = 21.82$, $p = 0.994$), these were not consistent with the predicted value of 14.3 per cent of a perfectly Bayesian actor.⁹² Despite this, neither experiment reflected the tendency of decision-makers to ignore base rates outright. Over half of the participants in both experiments identified *Probability* with that of the base rate (that is, 10 per cent). This suggests that effort was exerted to reach a sufficiently acceptable, but not necessarily optimal, judgement and highlights the fact that humans tend to act as satisfiers rather than maximisers.⁹³ Further analysis reveals that for both experiments *Probability* ranged from 10 per cent to 60 per cent suggesting that both individuating and base rate information were considered by participants in their final assessment. This allows the claims of **Hypothesis 1** to be challenged but surfaces a separate line of inquiry. If base rates are not neglected outright, then what factors cause the limited use of both the base rate and individuating information resulting in suboptimal judgements?

Information order and bias

For Experiment 1, *Order* is the sole factor influencing the participants' judgements. **Hypothesis 3** postulates that given cognitive constraints, individuals use the most recent information made available. This *recency effect* is well documented in the literature but appears to be absent in this case. Instead, the results suggest that earlier information plays a larger role in the inflating *Probability* and that the *primacy effect* is what is being observed instead.

The *primacy* and the *recency* effects illustrate the tendency of individuals to recall the first and last items in a series best. These are the consequences of limited cognitive resources. The *primacy effect* is believed to occur since initial information is more effectively stored and accessed in long-term memory. In contrast, the *recency effect* occurs due to the limited availability of short-term or working memory, which requires less effort to access. However, the *recency effect* is mitigated when additional tasks involve the use working memory.⁹⁴

Individuals can store approximately seven (7) items (plus or minus 2) in short-term memory at any given time.⁹⁵ With respect to the vignette, participants would have to take note of at least seven (7) details prior to the presentation of base rate and individuating information. This increases the likelihood of using earlier information (either of the above depending on *Order*) in determining probability due to the cognitive limits imposed.

Empirically, the fact that the value of *Probability* is higher when *Order* is reversed supports this claim. If the *recency effect* is present, then treatment groups where individuating information is

⁹²Since the vignette presented the base rate for states, in general, as the source of cyber operations; a perfectly Bayesian actor should indicate a value even lower than 14.3 per cent.

⁹³Herbert Simon, 'Invariants of human behavior', *Annual Review of Psychology*, 41:1 (1990), pp 1–20.

⁹⁴Robert Bjork and William Whitten, 'Recency-sensitive retrieval processes in long-term free recall', *Cognitive Psychology*, 6:2 (1974), pp. 173–89.

⁹⁵George Miller, 'The magical number seven, plus or minus two: Some limits on our capacity for processing information', *Psychological Review*, 63:2 (1956), p. 81.

presented last should have a higher *Probability* compared to when it is provided earlier. However, the results illustrate that *Probability* is, on average, 7.58 percentage points lower in these cases. Inversely, when base rates are presented last, *Probability* increases. This implies that participants access information provided earlier more readily than those received later. This may be due to the exhaustion of working memory and the need to retain as much information as possible.

Alternatively, another reason for this observation may be due in part to the structure of both the individuating and base rate information. Other experiments demonstrate that both format and length account for the decision to use or ignore specific pieces of information.⁹⁶ Longer formats or mathematically complex ones encourage subjects to choose simpler and easier to comprehend alternatives. This, however, cannot account for preceding observation as the vignette was structured in such a way that both individuating and base rate information are reasonably similar in terms of length and format. As such, it is unlikely that tendency to favour the initial information is a function of its appearance.

Similarly, the effects of a limited attention span may also be dismissed. The experiment employs attention checks meant to ensure that participants are focused on the task at hand. Those that fail this are removed from the study. Similarly, the experiment only involved participants that have been favourably rated by other researchers. While this does not necessarily guarantee their attention, it does suggest that a certain level of engagement with the material.

Given the above observations and the unlikely occurrence of alternative explanations, this weakens the proposition of **Hypothesis 3** regarding the *recency effect* but does not discount the importance of the order in which information is presented. This finding raises three important implications regarding the attribution of these events.

Procedurally, evidence in the wake of a cyber operation is not received or processed in a fixed order. Although Rid and Buchanan⁹⁷ have suggested the Q-Model in which tactical, operational, and strategic information are best processed in that sequence, they acknowledge that the process can start in any of these stages depending on the availability of information.

Organisationally, the continued shortage of cyber security expertise in government institutions has stimulated programmes that recruit from the private sector. These individuals enter these roles without the experience and, more importantly, without motivated biases that their counterparts from government may have. Methodologically, this aligns them more closely with the profile of students participating in this experiment and may manifest the same behaviours. As such, processes and technologies aimed at managing the volume of information are a worthwhile investment to avoid bias such as that encountered in Experiment 1.

Politically, and perhaps more importantly, if more recent information is treated as salient, actors involved in persistent disputes in either the physical or virtual domain are likely to ignore existing probabilities. As cyber operations are increasingly employed as adjunctive tools, these are likely to be framed in the context of the existing relationship between parties involved. Furthermore, a sudden escalation in the virtual domain may be perceived as an escalation of the conflict itself.⁹⁸ For instance, the disruption of the Ukrainian power grid was quickly attributed to the Russian regime as an escalation of the conflict when in fact subsequent analysis showed that these systems had been compromised prior to the conflict and may not have been linked to the desire to escalate.⁹⁹

And while the above case and subsequent research has yet to demonstrate the occurrence of cross-domain escalation, its absence does not discount its possibility in the future.

⁹⁶Gordon Pennycook, Dries Trippas, Simon Handley, and Valerie Thompson, 'Base rates: Both neglect and intuitive', *Journal of Experimental Psychology-Learning Memory and Cognition*, 40:2 (2014), pp. 544–54.

⁹⁷Rid and Buchanan, 'Attributing cyber attacks', pp. 4–37.

⁹⁸Gartzke and Lindsay, 'Thermonuclear cyberwar', pp. 37–48; Maness and Valeriano, 'The impact of cyber conflict on international interactions', pp. 301–23.

⁹⁹Kim Zetter, 'Inside the cunning, unprecedented hack of Ukraine's power grid', *Wired Magazine*, available at: {<https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/>} accessed 29 November 2018.

Consequently, the tendency to rely on earlier information increases the risk of misattribution, threat inflation, and possibly escalation in response to malicious events in cyberspace.

Enemy images and bias

Although it may be comforting to suggest that motivated bias may be done away with by recruiting outside government or military institutions, most national cyber security organisations are staffed by these individuals. As such, their pre-existing beliefs become a crucial component in judgement. The results of Experiment 2 demonstrate that the effect of *Images* in the context of cyber operations mimics that of other conflicts such that judgements are formed in accordance with pre-existing beliefs. *Image*, however, only appears to be of consequence with participants in the second experiment. Accounting for the difference between samples may be found in the fundamental characteristics of participants and the overarching theme presented in the vignette.

The second experiment involved noticeably older individuals who were, on average, at least nine years older than those in the first. This age gap suggests that the former may have a greater interest in politics than the latter as they may operate outside an academic environment and would have a greater stake in day-to-day affairs.¹⁰⁰ Complementing this difference in priorities, the theme presented in the vignette (that is, territorial disputes) may resonate more given its political, economic, and social implications. The attempt to seize territory could have increased the salience of the treatment with participants who are more attentive with respect to current events and are politically engaged. This observation does not necessarily raise questions regarding the validity of this experiment but instead highlights the possibility that exogenous factors can have a direct influence on experimental outcomes. In this case, it serves to reinforce the impact of the treatment used (that is, *Images*) and results in *Probability* increasing by 9.32 percentage points when participants are treated to an aggressive and uncooperative neighbour. As with the student sample, participants in this experiment appear to integrate both base rate and individuating information into their evaluation. *Order*, however, is not statistically significant.

This does not discredit **Hypothesis 2.1** but requires its assumed outcome to be relaxed from outright neglect to that of an inflated assessment. The importance of this finding cannot be disregarded given ongoing efforts to securitise the domain. Since 2007, the number of states with national cyber strategies has more than doubled. Furthermore, with states investing in offensive capabilities; the effect of casting cyberspace as a domain of conflict between rivals risks the possibility of a security dilemma and unintended escalation.

The nature of cyberspace finds parallels with the enduring challenges of ascertaining another state's capabilities.¹⁰¹ The limited lifespans of cyber weapons, to complicate matters, renders public demonstrations infeasible. Motivated by the fear of a 'bad actor's' intentions, states may engage in pre-emptive cyber operations to obtain intelligence regarding a rival's cyber capabilities. These operations require the compromise of a rival's cyber infrastructure to serve as either: (1) a channel through which sensitive information is exfiltrated; or (2) as an initial stage of a larger and (potentially) damaging operation. While the former is accepted as routine for interstate relations, the latter signals malicious intent that ought to be contained. Unfortunately, cyberspace is not conducive for signalling that allows for these to be distinguished from one another and may result in an unintended security dilemma.¹⁰²

The tools employed to exfiltrate information can be modified to disrupt or degrade crucial systems. Short of an initiator publicly declaring their intent, the discovery of these operations can, at

¹⁰⁰Hafner-Burton, Hughes, and Victor, 'The cognitive revolution', pp. 368–86; Mintz, Redd, and Vedlitz, 'Can we generalize ...', pp. 757–76.

¹⁰¹Borghard and Lonergan, 'The logic of coercion in cyberspace', pp. 464–6; James Fearon, 'Rationalist explanations for war', *International Organization*, 49:3 (1995), pp. 379–414.

¹⁰²Borghard and Lonergan, 'The logic of coercion in cyberspace', pp. 456–9; Buchanan, *The Cybersecurity Dilemma*, pp. 48–9.

worst, result in an escalatory spiral between states.¹⁰³ Should the victim of these pre-emptive operations harbour pre-existing beliefs regarding the malicious nature of the initiator, it is likely that the former would respond in a manner that reinforces this belief.

Given the absence of norms regulating action in cyberspace, the influence of pre-existing beliefs suggests that a state's response is informed by its experience with a suspected aggressor. To minimise losses, states faced with a consistently belligerent adversary may opt to develop offensive capabilities that serve to deter further attacks or to respond in kind. Consequently, the suspected aggressor may increase its capabilities in cyberspace resulting in an arms race between these states. While most scholars are sceptical as to the extent of damage possible through cyber means alone, this escalation does not serve to ease tensions in an already unstable and hostile cyberspace.

The influence of role

A curious outcome of this study is the seeming insignificance of *Role* for both experiments. The influence of organisational role and accountability is well documented in the literature. As a factor that aggravates or mitigates bias, its muted effects are surprising to say the least. For both experiments, the lack of significance appears to be confirmed by its small effect size and limited statistical power ($f^1 = 0.046$, $1 - \beta_1 = 0.091$ and $f^2 = 0.006$, $1 - \beta_2 = 0.051$). These observations lead to three possible explanations to account for this outcome.

One is the inability of the treatment to elicit the desired affect from the participants. Given that both samples are unlikely to have ever found themselves in a position of authority, it is possible that simply stating their role in this vignette as the administrator of the NCC did not form a substantive affective association with the possible consequences of tarrying or rushing into a decision. A possible work-around for future experiments is to engage in power priming participants. Galinsky *et al.*¹⁰⁴ demonstrate that an individual not normally in a position of authority could be primed to act in such a way as if they were through specific recall tasks. Inversely, individuals could also be primed to behave as if they were in not in a position of authority.

A second possibility that complements the first is the absence of consequences associated with the failure to adhere to the expectations of a participant's assigned role. As was stated in the preceding theoretical framework, costs are often associated with positions of greater responsibility. Failure to act accordingly may result in punishments stemming from organisational/bureaucratic practices or from the electorate (for example, not getting re-elected). Experimentally, these consequences may be simulated through graduated rewards based on performance. For this study, the use of a uniform reward system may have omitted this consequence-based reasoning from the participants resulting in *Role* being statistically insignificant.

A third possibility is that cultural factors may contribute directly to participant behaviour. Despite efforts to increase variation using Internet-based experiments, participants were predominantly from the United States and Europe. Although variations in behaviour as a function of culture is by no means novel, the results of the previous experiments do not allow for this assumption to be tested.

While the first two possibilities may be addressed through a redesign of the experiment to use priming and a graduated reward system, administrative and logistical considerations¹⁰⁵ rendered this infeasible at the time. Instead, a third experiment was conducted using Amazon's Mechanical Turk (*MTurk*) to account for cultural differences. The decision to perform this on *MTurk* was due to its demographic differences with *Prolific*; specifically its larger number of Asian participants.

¹⁰³Gartzke and Lindsay, 'Thermonuclear cyberwar', p. 45.

¹⁰⁴Adam Galinsky, Deborah Gruenfeld, and Joe Magee, 'From power to action', *Journal of Personality and Social Psychology*, 85:3 (2003), p. 453; Adam Galinsky, Joe Magee, Ena Inesi, and Deborah Gruenfeld, 'Power and perspectives not taken', *Psychological Science*, 17:12 (2006), pp. 1068–74.

¹⁰⁵This would have required the experiment to be re-registered and for additional funding sought.

A total of 120 participants with 15 participants per treatment group were recruited for this experiment. Of these, 25.83 per cent (31) were from Asia. When modelled to include the dummy variable *Region*, only *Role* ($f = 0.310$, $p = 0.001$) and *Region* ($f = 0.274$, $p = 0.017$) were found to be statistically significant at the $p < 0.05$ level.¹⁰⁶ *Probability* increased by 14.33 percentage points when participants were from Asia compared to those from the Americas. An increase of 15.04 percentage points was also observed when assigned an administrative role. There is, however, no significant interaction between *Region* and *Role*. But this is most likely due to a lack of statistical power for this experiment ($f = 0.144$, $1 - \beta = 0.255$).

This follow-up experiment suggests that cultural factors may moderate the impact of *Role* in the formation of judgement. Given the relatively small sample size, however, confidence in this outcome is in question. Moreover, the importance of having an affective association with a given role and the threat of consequences cannot, as of yet, be ruled out. Consequently, these findings do not allow us to confirm or refute **Hypothesis 2.2**.

Theoretical, methodological, and policy implications

The result of the experiments reflect the apprehension associated with degradative cyber operations. While base rates were not found to have been dismissed outright, judgements tend to inflate the likelihood of state involvement resulting in suboptimal judgements such as misattribution. The complexity that characterises cyberspace encourages decision-makers to resort to heuristics to reduce the cognitive load required. Although this permits faster processing, it increases the likelihood that only a handful of facts are considered, resulting in biased judgements. In the experiment, the appearance of higher probabilities earlier in the vignette prompted participants to base their assessments on these values resulting in a stronger belief that Country B is indeed the aggressor. Similarly, the context that surrounds these events encourages the maintenance of pre-existing beliefs. When informed of State B's past aggression, decision-makers perceived a greater probability of their involvement. This serves to reinforce the idea that Country B is indeed a bad actor willing to further their own interests.

Theoretically, these findings are neither novel nor surprising. The role of both heuristics and motivated reasoning in political decision-making has been extensively studied. These concepts, however, remain understudied *vis-à-vis* cyber operations. Consequently, the results contained herein are a decisive development within the field for two reasons. First, they serve to dispel the myth of the domain's exceptionalism. While there is no denying that the manufactured nature of cyberspace enables actors to behave in ways different from the conventional domains of air, land, sea, and space; judgements are ultimately the product of human cognition. As such, these outcomes are subject to cognitive limitations – motivated or otherwise – that manifest themselves across operational domains.

Second, this behavioural shift in the study of interstate interactions in cyberspace complements the ongoing efforts at the state and systemic levels. Despite its emergent status, the cyber security literature is rich with studies that leverage existing state-level theories to explain the use of cyber operations. Recently, this has served to temper persistent belief in the revolutionary potential of these operations but fails to penetrate the black box surrounding decision-making processes. Experiments such as this extend the findings of earlier efforts to better evaluate the behaviour of individuals that make up the state and are ultimately tasked to respond to threats and opportunities within the domain.


Methodologically, the study imparts important lessons for future researchers. Since its earliest days, the field of cyber security has relied on observational studies to both propose and develop appropriate theoretical frameworks. The opaque nature of this phenomenon, unfortunately, results in limited, and possibly, biased data. Consequently, this study demonstrates the feasibility of an

¹⁰⁶Experiments 1 and 2 were also tested to include *Region* but no significant results were found.

experimental design in appraising this phenomenon. However, two caveats exist in this respect. First is the importance of selecting appropriate samples. While undergraduate students are easily accessible, their knowledge and experience may prove limited as argued by other scholars. Although this has not negatively impacted this study, this is not always the case. Therefore, careful participant selection is needed prior to the start of any experiment. Second is the recognition and study of organisational and bureaucratic processes associated with foreign policy. The study did not tackle this aspect and it is prudent to point out the policies are not the sole creation of an individual. Even the most despotic of regimes depend upon a small circle of advisers. While the observed biases are unlikely to completely disappear when decisions emerge from group interactions, these may either be tempered or aggravated in kind. As such, future studies would do well to develop more complex experimental designs that better simulate the distributed and shared nature of decision-making.

The results of the experiments also highlight the continued need for technologies and processes to assist in the analysis of security incidents. Even the best-trained analysts are constrained by their cognitive abilities. It is then appropriate to invest in tools that reduce this workload that, in turn, minimises dependence on fast – but possibly flawed – cognitive short-cuts. Similarly, scholars such as Rid and Buchanan¹⁰⁷ have called for processes that divide analysis between groups and individuals best suited to analyse information at different points in the analytical process. Building on this argument, national cybersecurity organisations may also benefit from hiring individuals from a diversified pool. An increase in variation may serve to temper biases prevalent in one group compared to another.

While the outcome of this study serves only as an initial foray into a larger research programme, its significance should not be judge as trivial. Its theoretical, methodological, and policy implications validate the need to broaden our study of cyber security to include individual behaviour as a crucial aspect of this domain.

Author ORCIDs.  Miguel Alberto N. Gomez, 0000-0001-8575-2188

Acknowledgements. I would like to extend my gratitude to my colleagues at the Center for Security Studies at the ETH for their initial comments, which helped to set the tone of this article. I would also like to acknowledge my reviewers, whose comments serve to strengthen the arguments presented herein.

Miguel Alberto Gomez is a Senior Researcher at the Center for Security Studies at the ETH in Zurich. His research explores the role of heuristics and cognitive bias in the formation of cybersecurity policy and interstate interactions in cyberspace.

Appendix

Vignette structure

For the past ten years, a territorial dispute has existed over a series of islands in the sea that separates your country from your southern neighbour. Although unpopulated, maritime surveys have discovered rich oil deposits in the area. Your country's claim is based on the location of these islands within your territorial waters. In contrast, your southern neighbour insists that it has an underlying historical claim represented in a series of antiquated documents. Two years ago, both countries presented their respective cases to the International Court of Arbitration in The Hague where the court confirmed your country's legal right to the territory. In response, your country has begun to deploy offshore rigs to harvest these oil deposits.

Image

Non-Enemy

In previous territorial disputes, both you and your southern neighbour have cooperated with one another. Despite your respective self-interests, both are motivated by the need to maintain peace.

¹⁰⁷Rid and Buchanan, 'Attributing cyber attacks', pp. 4–37.

Enemy

In previous territorial disputes, your southern neighbour deployed its military to occupy the contested territory. Despite efforts to ensure peace, your southern neighbour appears to be motivated primarily by self-interest.

Yesterday, a cyber attack targeted your country's off-shore rigs in the disputed territory. The attack appears to have employed advanced tools and techniques and disrupted operations on these off-shore rigs for several hours.

Role

Non-Administrative

As one of many analysts with your country's National Cyber Command, you are tasked with evaluating the available evidence to provide your superiors with information regarding cyber attacks against your country.

Administrative

As the head of your country's National Cyber Command, you are tasked with providing the head of state with information regarding cyber attacks against your country that will be used to formulate the ensuing response.

Order

Original

Trend data indicate that 10 per cent of cyber attacks are initiated by foreign governments while the remaining 90 per cent originate from non-governmental sources (for example, criminal organisations, private individuals, etc.). Although the investigation is still underway, those involved have begun the process of identifying the source of the attack. Early reports concerning the source of cyber attacks are correct 60 per cent of the time but are incorrect for the remaining 40 per cent.

Reversed

Although the investigation is still underway, those involved have begun the process of identifying the source of the attack. Early reports concerning the source of cyber attacks are correct 60 per cent of the time but are incorrect for the remaining 40 per cent. Trend data indicate that 10 per cent of cyber attacks are initiated by foreign governments while the remaining 90 per cent originate from non-governmental sources (for example, criminal organisations, private individuals, etc.).

As information continues to be collected, you have been asked to provide an initial assessment of the situation.