JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2019, 1–19
doi: 10.1093/cybsec/tyz012
Research paper

Research paper

Past behavior and future judgements: seizing and freezing in response to cyber operations

Miguel Alberto Gomez *

Center for Security Studies, ETH Zurich, Haldeneggsteig 4, Zuerich, 8092, Switzerland

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

Received 22 April 2019; revised 25 June 2019; accepted 27 August 2019

Abstract

The use of cyber operations as a foreign policy instrument continues to stimulate academic interest towards interstate behavior in this domain. With continued investment in offensive cyber capabilities, there is an urgency to provide both academics and policy-makers with a better grasp of this phenomenon. While the past decade saw the growth of frameworks that highlight systemic and/or technological factors, this article investigates the role of pre-existing beliefs in the attribution of malicious cyber operations. Through survey experiments, it highlights the phenomenon of *seizing* and *freezing* with respect to attributive judgements in response to degradative cyber operations. With respect to theory, the results contribute to the emerging study of the cognitive–affective aspects of cyberspace. As for policy, the results illustrate the potential for biased judgements in response to incidents and reinforces the need to develop mechanisms that minimize its impact on state behavior.

Key words: bias; heuristics; attribution; psychology; survey experiments; elites

Introduction

To what extent do foreign policy elites base their evaluation of cyber operations on established behavior? Although the exercise of cyber power is perceived by some as a novel development in interstate relations, the context that colors these interactions are not. Interactions in cyberspace are not a-strategic and are framed by larger strategic and political objectives [1]. From territorial disputes to questions of regime legitimacy, cyber operations complement conventional capabilities that are at the disposal of capable and motivated actors. As such, these are subject to cognitive constraints that frame state interactions in physical space. Consequently, this article argues that foreign policy elites are equally prone to the phenomena of *seizing* and *freezing* due to pre-existing beliefs concerning adversaries when attributing malicious behavior in cyberspace.

Seizing reflects a pre-disposition to gravitate towards cues that appear to confirm pre-existing belief(s). *Freezing*, in contrast, is the refusal to adjust existing judgements due to the need to maintain beliefs[1] [2]. Both are linked to the human need to maintain cognitive consistency in a complex and uncertain environment and is further aggravated by our predisposition for closure.[2] And while both have been studied in response to behavior within the physical domain, their significance in the context of interstate cyber interactions remains understudied [3–6].

While the preceding statement suggests that the article does not provide a novel contribution, such an assessment is myopic if not viewed in the wider context that colors interactions within this domain. First, malicious behavior in cyberspace often—if not exclusively—involves established rivals [7]. Second, certain actors, such as the USA, are not only investing in increased capabilities in cyberspace; but are practicing doctrine that espouses the increased exercise of cyber power (i.e. persistent engagement). Although adversarial interactions in cyberspace introduces a degree of strategic stability,[3] an unexpected shift in these interactions by the appearance of cyber operations exhibiting greater damage potential may facilitate an imbalance in the existing relationship [8]. And while the target may successfully contain this emergent threat, expectations built on an adversary's past behavior may motivate a

1 Both seizing and freezing can be viewed as an attempt to avoid cognitive dissonance.

2 Prevalent in high stress scenarios such as interstate disputes.

3 Rivalry relationships in cyberspace are thought to be fairly stable due to pre-held notions of how each party is expected to be have and the extent of their aggression.

disproportionate response to pre-empt any further threats [9]. This possible destabilization facilitated through misperception rooted in uncertainty inherent in the cyberspace and pre-existing beliefs reinforced by past interactions is rarely acknowledged by the literature that often assumes a uniform perception of the nature and threats to and from this domain [10].

This article, in response, employs Internet-based survey experiments to better understand how attributive judgements form in an environment colored by adversarial expectations and technological uncertainty. By establishing the presence (or absence) of *seizing* and *freezing*, it highlights the manifestation of biased judgements that may facilitate an escalatory spiral between rivals. This line of inquiry is crucial in terms of the formation of adequate foreign policy that tackles the growing exercise of cyber power. Consequently, the findings provide noteworthy theoretical, methodological, and policy contributions to the field of cyber security.

With this in mind, the article is divided into four main sections. The first provides the reader with the theoretical underpinnings of the article. Specifically, it surfaces the mechanisms that link judgements with pre-existing beliefs. This is then succeeded by a discussion of the experimental design employed and its limitations. Readers should be made aware that it is not the goal of this article to generalize its findings. Rather, it serves to demonstrate the internal mechanisms at work through which foreign policy elites may be subject to under the specified conditions. Once the design is explained, the results of the experiment are summarized. Finally, the remainder of the article provides a discussion of how the results serve to reinforce the proposed framework and its implications for policy.

Theoretical framework

Framing cyber operations

Cyber operations are defined throughout as the exercise of power through cyberspace as a means of achieving a state's strategic objective. The presence of a cyber operation signifies strategic intent behind the use of this instrument. The emphasis on strategic utility serves to exclude actors such as private individuals or independent criminal organizations whose actions may not necessarily have any discernable strategic goals. This does not, however, suggest that non-state actors are inconsequential. Rather, those whose actions are not aligned with a state's strategic objectives are not of particular interest for this article.

While an agreed upon taxonomy of cyber operations has yet to emerge, events in cyberspace are broadly categorized as either espionage, disruptive, or degradative [11]. Of these, the article concerns itself with degradative cyber operations.

These operations reflect an aggressor's technical sophistication and organizational maturity through its latent potential to attain strategic goals by means of inflicting harm by risking an adversary's cyber assets. These exploit vulnerabilities that are typically beyond the reach of actors without the necessary competencies.[4] And while cyberspace is still perceived to reduce material imbalances, previous incidents prove otherwise. Operational outcomes depend not only on technical prowess but also on the ability to harness other policy instruments and complemented by organizational maturity[5] [12–14].

The strategic potential of degradative operations is further rooted in the underlying nature of cyberspace. The domain is composed of a multitude of interconnected and interdependent systems [15] that increase complexity and limit the identification and rectification of flaws [16]. This situation of unknowability and vulnerability is further aggravated by the possibility of discovery and exploitation by a malicious actor [17]. As states become increasingly reliant on Information and Communication Technologies (ICT), an adversary's ability to impose risk by cyber means results in a feeling of helplessness and inevitability. Moreover, the interconnectedness between components of cyberspace implies that operational consequences may cascade across different components of the domain, increasing the potential for damage [18].

This uncertainty in conjunction with the potential consequences of these operations increases dread risk among those affected by it [19]. Defined as risk associated with high-impact low-probability events, this furthers the likelihood of sub-optimal judgements that emerge in these situations [20, 21]. Reinhardt [22] notes that individuals without direct experience of these events display stronger reactions to it. This exaggerated response towards unlikely events is crucial with respect to degradative operations. Based on available data, degradative operations constitute approximately 13% of all cyber operations from 2000–16 [23]. This stands in stark contrast with media reports over a comparable period that suggest not only the greater frequency of these events, but also an increased potential for damage [24]. Furthermore, given the strategic context in which these occur, individuals perceive these events through the lens of history. For instance, should Country A's power plants be disrupted via cyber means; an ongoing dispute with Country B may enable the assumption that the latter is responsible for this incident. This belief may persist despite evidence to the contrary if Country B has been consistent in his behavior from the point-of-view of an observer.

This is not to say that less damaging events such as disruptive or espionage-type operations are inconsequential. However, the focus of both policy elites and the media on these low-probability but highly damaging events merits inquiry. This is further necessitated by the now common understanding that events in cyberspace occur below the threshold of armed conflict and sustained exchanges within the domain are growing more frequent. With the standard regulating the exercise of cyber power relaxed, states are increasingly developing technological and organizational capabilities required for complex, and consequential, operations. With these in mind, the article focuses on degradative cyber operations as the phenomenon of interest as these are most likely to elicit sub-optimal judgements.

A dearth of cognitive frameworks

Existing frameworks that explain interactions in cyberspace presuppose that both aggressors and targets are unitary rational actors. Unitary in the sense that state behavior is treated as the result of a single monolithic entity. While rationality requires the evaluation of alternatives and consequences prior to a decision. This is problematic as it ignores endemic individual-level attributes and organizational constraints[6] [4, 25–27].

Although rationality is assumed in most structural theories, cognitive idiosyncrasies either strengthen or weaken this assumption. Instead of an objective evaluation of the situation, individuals

4 In the form of technical expertise, financial resources, and organizational maturity.

5 Reflected through the ability to plan, execute, and maintain complex operations.

6 While this is a crucial aspect in the formulation and execution of policy, this feature is beyond the scope of this study and is to be pursued by the other at a later date.

routinely employ cognitive shortcuts and/or motivated reasoning when formulating judgements under conditions of uncertainty [28–31]. This does not imply that foreign policy elites are irrational individuals. They are instead constrained by a host of internal (cognitive) and external (systemic) factors resulting in bounded rationality and satisficing behavior [32]. With both the complexities of interstate relations and the expertise required to understand cyberspace, evaluating every possible outcome is cognitively infeasible and invites the use of heuristic mechanisms. And while heuristic usage may result in fast and accurate judgements, these can just as easily encourage misperceptions that manifest as inappropriate foreign policies [33, 34].

Besides cognitive limitations, sub-optimal judgements also occur as a function of pre-existing beliefs. Beliefs serve to simplify a complex and uncertain environment that, in turn, enable certain cognitive processes [3, 35]. These include influencing and setting expectations, the plausibility of certain propositions, the reinforcement of said propositions, and the need to mitigate cognitive dissonance [35]. Taken collectively, it follows that the presence of certain beliefs limit the integration of new information that may possibly contradict these preconceived notions.

This deviation from rationality is made especially salient during periods of uncertainty—a feature common in cyberspace [19]. This, in turn, increases the risk of a security dilemma unfolding. Cyberspace is not exempt from this phenomenon due to limitations in assessing an adversary's capabilities and intent [12]. This problem, however, is compounded by the unique technological and the socio-political milieu that surrounds cyber interactions.

The introduction of new technologies is thought to modulate the intensity of the security dilemma [8, 18, 36]. When adversaries are unable to distinguish between offensive or defensive capabilities, the need for security may result in a security dilemma. Within the physical domain, communication and capability demonstrations serve to reduce the potential for confrontation. Demonstrating capabilities in cyberspace, however, risks limiting the opportunities of employing these at a later time [37, 38]. Burning cyber tools as a means of signaling offers adversaries the opportunity to develop countermeasures. Moreover, seemingly benign tools used for espionage may also serve as a stepping stone to potentially more damaging operations as the discovery of espionage tools are not an adequate gauge of intent [8].

The socio-political aspects of cyberspace also serve as a source of uncertainty. Differences in domain conceptualization results in a variance of threat perceptions across actors. This inconsistency in perception may lead an adversary to underestimate the perceived cost of their action resulting in unintended escalation [39, 40]. Given the nature of disputes that color exchanges in cyberspace, such a misestimation increases the risk of unintended escalation [7].

Maintaining beliefs

As previously stated, rationality requires the identification of a solution once all possible alternatives and outcomes have been evaluated. Previous responses to cyber operations do not appear to conform to this standard. In 2007, Estonian officials held firm to the belief in the Russian government's culpability despite evidence suggesting otherwise [41]. Similarly, the recent incident during the Pyeongchang games saw the misattribution of the incident to North Korea [42]. Much earlier, the USA misattributed the 1998 Solar Sunrise incident to Iraqi operatives given the underlying tensions at the time [43].

Researchers note that the availability of new information serve to either re-affirm pre-existing beliefs if congruent or are ignored if contradictory [3, 6, 35, 44–47]. Although beliefs may be challenged and replaced in the face of irrefutable information, such a situation is rare in the context of interstate relations [48]. Beliefs operate as filter through which other perceptual processes occur [49, 50]. Of interest for this article is the emergence of motivated reasoning as a means of maintaining these beliefs in the presence of contradictory information.

Researchers posit that motivated reasoning depends on past affect-laden experiences. When presented with new information, individuals draw from affect-laden information stored in long-term memory. Once accessed, a specific affect is triggered along with the associated information. A heuristic mechanism then evaluates one's feelings with respect to the information activated and reinforces the existing affect regardless of the presence of dis-confirmatory evidence [51, 52]. Relatedly, Mercer [45] asserts that emotions function as an assimilative mechanism for beliefs. Whereas rationalists assume that new information serves to update existing beliefs that eventually converges towards reality, Mercer instead argues that emotions assimilate data into beliefs—thus reinforcing it. These support Jervis' [3] rationale for maintaining beliefs and provides the necessary psychological basis for this argument. These beliefs are reflected as 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" [53]. Specifically, the article is interested in the construct of enemy images through which other actors are perceived to behave in "bad faith" [54, 55]. While these images are ingrained for a number of reasons, the article identifies conflict accumulation as the primary reinforcement mechanism. Dreyer [56] notes that constant exposure to multiple issues in a rivalry environment reinforces these images and increases the likelihood of conflict. Crucial to this notion is the idea that an enemy image is strengthened irrespective of subsequent issues. Consequently, succeeding "negative" interactions strengthen pre-existing images. Individuals employ "gist" when retrieving information from long-term memory; relying on salient, rather than specific, features when drawing parallels between events at different points in time [57].

The need to maintain beliefs is associated with the desire for cognitive closure [58]. For foreign policy elites, associated costs are the function of role and accountability [59]. This is particularly true in democratic regimes where elites are held accountable for policy failures. Self-interest motivates elites to maintain existing beliefs rather than invest in time and cognitive resources to consider alternatives [4]. Consequently, closure as a function of role occurs quickly (urgency) and must be held constant (permanence) [60].

Urgency motivates individuals to "seize" on 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" on their judgements. As such:

> ***Hypothesis 1.*** Decision-makers are likely to seize on evidence that confirms pre-existing beliefs.
>
> ***Hypothesis 2.*** Decision-makers are likely to freeze on evidence to maintain the stability of beliefs.

Although the literature demonstrates that enemy images significantly influence judgements in the context of cyber operations, these do not focus explicitly on the tendency of decision-makers to seek out confirmatory evidence nor do these look into the extent to which dis-confirmatory evidence was dismissed [61]. Consequently, this

article serves to extend previous findings so as to better understand judgements in response to malicious cyber incidents.

Experimental design

Treatment

The experiment is undertaken using an Internet-based between-groups survey experiment.[7] Participants are presented with a vignette in which their country of Idimore[8] is experiencing a cyber operation that continues to disrupt the ongoing national elections.[9] A subsequent investigation reveals that their rival country of Vadare may have been the source.

The treatment is induced through the manipulation of a pre-existing belief (*Enemy*) across two levels. Participants exposed to the treatment are informed of Vadare's past belligerent behavior towards their country. Other participants are told nothing of the suspected aggressor's past actions and serves as the control group. After reading the vignette, participants are asked how confident (*Confidence*) they are that Vadare is the aggressor. They are required to provide an explanation to justify their reported level of confidence. If *Hypothesis 1* is correct, participants exposed to the treatment will seize on evidence confirming Vadare's responsibility and would be reflected both in the reported *Confidence* level (higher) and justification.

Participants are then presented with new information that casts doubt on Vadare's culpability. This surfaces the phenomenon of *freezing* given the preceding treatment. Participants are again asked to indicate their level of confidence regarding the role of Vadare and are again asked to explain their judgement. If *Hypothesis 2* is correct, the difference between the current and previously reported level of confidence (*Confidence Shift*) for those in the treatment group should be larger in magnitude and positive in direction compared to the control. Additionally, the provided justification should be aligned with the enemy image provided through the treatment.

To rule out competing explanations, attributes such as *Risk Tolerance*, *Cyber Event Awareness*, and *Exposure to Cyber Crime*, are measured [62]. *Risk Tolerance* is considered since risk acceptant individuals are more likely to form judgments aligned with their beliefs despite the absence of more information [63]. On the other hand, *Cyber Event Awareness* and *Exposure to Cyber Crime* are included as these may serve as reference points on which to base judgements and to complement to the information provided.

Since the experiment is distributed via the Internet, steps are taken to safeguard the quality of data collected. Two (2) attention check questions are included at different stages. Participants that fail both are removed from the experiment. As an additional step, a *t*-test is conducted on the reported level of confidence prior to and immediately following the presentation of new information to determine whether the desired effect was achieved. Participants are instructed to refrain from referring to external sources of information such as the Internet. While not guaranteeing compliance, studies have shown the participants do adhere to these instructions for the most part [64, 65].

Unit of analysis

The unit of analysis is the individual. Although organizational and strategic considerations also play a role,[10] these are currently out-of-scope. Given that cultural traits may influence behavior, participants are recruited using the platform *Prolific Academic* to maximize variations in this regard. It is also important to note that the behavior of Internet-based participants do not significantly differ from their lab-based counterparts and would allow for reasonable comparisons to be drawn [64, 65].

Assuming a moderate treatment effect ($f^2 = 0.15$) significant at the 0.05 level and with a power of 0.8, pre-test power computations indicate a minimum of approximately 80 samples are required.

Although most experimental studies in Political Science and International Relations employ undergraduate students, criticisms regarding external validity have been raised[11] [66, 67]. To mitigate these challenges, participants are primed into a position of power prior to reading the vignette. This priming requires them to recall an instance in which they were in a position of authority over an individual(s). Recall tasks encourage participants to behave and reflect characteristics of those in a position of power [68–70].

It is worth noting that the article concerns itself with foreign policy decision-makers. That is, individuals whose behavior the experiment attempts to replicate are those directly responsible for setting the direction of foreign policy rather that first responders to this issue. Consequently, this does not take into consideration military elites who may have a more competitive perception of the domain as well as the general population who do not have a direct role in the formation of foreign policy. By recruiting non-students (i.e. older individuals with work experience); it is hoped that participants possess a better understanding of politics, are invested in the outcome of specific governmental policies, and have experienced some form of authority (if not then the priming prior to experiment should address this concern). With respect to excluding military elites, previous experience demonstrates that these individuals are significantly underrepresented (<1%) within the *Prolific Academic* participant pool. Consequently, the chances of these individuals joining in the experiment is quite low.

Measurement and analysis[12]

The effect of the treatment is reflected primarily through the *Confidence* and *Confidence Shift* variables.[13] As previously discussed, the former represents a participant's reported level of confidence after reading the vignette. *Seizing* presents itself if the mean of *Confidence* for the treatment group is higher than that of the control. The latter, in contrast, is the different between the *Confidence* value immediately following and just prior to the presentation of new information. Effectively, *Confidence* Shift is computed as follows: $Confidence_{Post} - Confidence_{Pre}$. *Freezing* presents itself if the mean of *Confidence Shift* for the treatment group is larger and positive in comparison to that of the control.[14]

Apart from these two dependent variables, two additional mediating variables are also measured for the purpose of this

7 The complete details of the experiment are available in the Appendix A. These may be used for replication purposes.

8 Fictitious countries are employed to minimize the possibility of bias if actual countries are used in the vignettes.

9 See Appendix A.1 for further reference.

10 Tackled in two other research projects.

11 Past research also demonstrates a marked difference in how different populations respond to incidents in cyberspace in an experimental environment, all else being equal [19, 16].

12 For replication purposes, the analysis performed and the corresponding data are available as a GIT repository via the following address https://github.com/mgomezPH/seize_freeze.git.

13 See Appendix A.2 for further reference.

14 Basically, when $Confidence_{Post}$ is larger than or equal to $Confidence_{Pre.}$

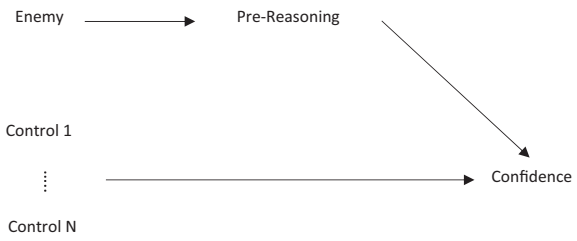

Figure 1. Confidence causal path.

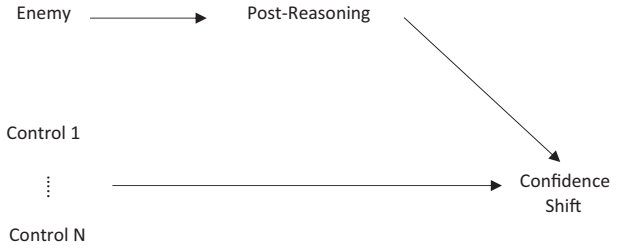

Figure 2. Confidence shift causal path.

experiment: *Pre-Reasoning* and *Post-Reasoning*.[15] *Pre-Reasoning* represents the justification provided by the participants after reading the initial vignette. This variable is derived from the open-ended response by identifying the presence (or absence) of confirmatory or disconfirmatory arguments. The former reinforces the belief in Vadare's culpability while the latter calls into question its involvement. The *Post-Reasoning* variable is coded in a similar fashion to capture variations in the participants' reasoning after being presented with new information that possibly contradicts those previously provided.

To quantify[16] both *Pre-Reasoning* and *Post-Reasoning*, the number of confirmatory arguments is measured against the sum of arguments made both in favor and against the proposed attribution. This generates a value between 0.0 to 1.0. If *seizing* and *freezing* are present, this value should be above 0.5 (i.e. balanced judgement) for participants in the treatment group.

Although it would suffice to simply utilize *Confidence* and *Confidence Shift* to test the above framework, the inclusion of *Pre/Post-Reasoning* serves to better replicate the process of judgement formation. That is to say, images trigger a deliberative process grounded in pre-existing beliefs that frame judgements which influence the reported levels of confidence. To demonstrate this mechanism, both *Pre/Post-Reasoning* are regressed on the treatment (*Enemy*) and the other control variables of theoretical relevance. This causal mechanism is illustrated in Figures 1 and 2 above.

Other than the treatment and mediating variables, other control variables are also considered.[17] *Risk Tolerance* is measured using the instrument developed by Kam and Simas [71]. This is computed by taking the mean of the answers provided by participants on a 7-item questionnaire.[18] *Risk Tolerance* scores less than 0.5 suggest risk-aversion while those above 0.5 indicate risk-acceptance. *Cyber Event Awareness*, measures a participant's knowledge of recent events in cyberspace[19] [62]. Using a 6-item questionnaire, participants are asked whether the following statements are True or False with the percentage of correct responses serving as the indicator. Values less than 0.5 suggests limited knowledge of these events while those above 0.5 indicate knowledgeability. Finally, *Cyber Crime Experience* measures the extent to which participants have been victimized through cyberspace[20] [62]. Using a 3-item questionnaire, participants are asked whether they have experienced cybercrime. Questions in which answers are in the affirmative are summed and divided by three, yielding an indicator ranging from 0.0 to 1.0. Values closer to 1.0 suggest a high level of victimization.

Results and analysis

Summary

A total of 191 participants were recruited for this experiment. However, attention checks and failure to follow instructions led to the removal of 15 participants resulting in a sample size of 176.[21] Furthermore, the sample was balanced between treatment groups by randomly selecting participants based on the size of the smallest treatment group.[22] The final dataset consisted of 164 participants with a mean age of 38.4 years. Participants are predominantly European (70.5%) with the remainder coming from North America or Australia.[23] Of these, 107 (60.8%) are female while 69 (39.2%) are male.

In terms of *Risk Tolerance*, participants appear to be risk-averse; the sample having a mean of 0.435. With regards to *Cyber Event Awareness*, 0.748 indicates a familiarity with events in cyberspace that are related to the phenomenon of interest.[24] Finally, a mean of 0.441 for *Exposure to Cyber Crime* indicates that participants have been victimized through cyberspace in one form or another.

With respect to the outcome variables, the treatment group attained a mean score of 7.707 for *Confidence* while the control scored 7.159. Although not significantly differing in value, a *t*-test confirms a statistical difference between the two groups ($P = 0.036$). As for *Confidence Shift*, the treatment group attained a mean score of −2.256 while the control was found to be − 1.720. These two groups were not found to be statistically significant at the 0.05 level.

15 See Appendix A.2 and A.6 for coding, computation, and interpretation.

16 As an alternative analytical approach, Structured Topic Modelling (STM) may be applied to determine how the respective treatment affects the emergence of themes utilized by the participants. This approach was tested on the data but the relatively small sample size (< 200) failed to fully utilize the advantages of this approach. Future replication studies may want to employ this as a means of addressing the potential biases associated with manual coding.

17 The corresponding Appendix discuss how these controls are obtained, computed, and interpreted.

18 See Appendix A.3 for further computation and interpretation.

19 See Appendix A.4 for further computation and interpretation.

20 See Appendix A.5 for further computation and interpretation.

21 To determine whether the new information after the vignette took effect (e.g. increasing uncertainty) a *t*-test is conducted on the reported levels of confidence. For this experiment, the mean difference of 2.789 was found to be statistically significant at $P = 0.000$.

22 See Table A1 in Appendix A.7

23 This over-representation is not surprising given that Prolific is a European research platform initially developed at Oxford University. Other platforms such as Amazon MTurk reflect a bias distribution in favor of certain populations; to better control for variance resulting from individual nationalities, participants were grouped based on their country's membership in the Five Eyes alliance. The Five Eyes is an Anglo-American intelligence-sharing alliance that share common views with respect to cybersecurity.

24 This is not particularly surprising since participation in Internet-based experiments suggests familiarity with technology and the issues surrounding it.

As for the mediating variables, a mean of 0.739 for *Pre-Reasoning* while those in the control group scored 0.563. A *t*-test confirms a statistical difference between the two ($P = 0.003$). *Post-Reasoning*, in contrast, has a mean of 0.546 for the treatment group while the control scored 0.423. A *t*-test did not find a statistically significant difference between these two groups at the 0.05 level.

In discussing the effects of *Enemy*, the analysis is divided into three stages. The first[25] investigates the effect of *Enemy* on the outcome variables without considering the mediators (*Pre/Post-Reason*), the second[26] takes the effects of these into account, while the third[27] surfaces the effects of *Enemy* on *Pre/Post-Reason*. Take note that the previously mentioned control variables are also taken into account during the analysis across these three stages.

Without taking *Pre-Reasoning* into account, exposure to *Enemy* is statistically significant and increases *Confidence* by approximately 0.56 across the different model specifications. Both *Cyber Crime Experience* and *Age* also contribute to increasing the overall confidence of the participant by approximately 1.00 and 0.03, respectively. In response to new information, neither the treatment nor any of the other control variables were found to have a statistically significant effect on *Confidence Shift*.

Once the mediating variables are introduced, a crucial change is noted in the models. When specified to include *Pre-Reasoning*, *Exposure* ceases to have a significant effect on *Confidence*. *Pre-Reasoning*, however, increases *Confidence* by approximately 2.2 and is statistically significant across the different specifications. As before, both *Cyber Crime Experience* and *Age* exhibit similar effects that remain significant. In response to new information, however, *Enemy* again becomes significant and decreases *Confidence Shift* by 0.8 across specifications. Contrasting this, *Post-Reasoning* increases *Confidence Shift* by approximately 2.0 and is statistically significant.

The observed difference between models that include *Pre/Post-Reasoning* suggests the mediating effects of these variables. Analysis reveals that *Pre-Reasoning* is influenced by *Enemy* which increases its value by approximately 0.17. This suggests that providing participants with enemy images increases arguments favoring the culpability of the perceived adversary by approximately 17%. Contrasting this, *Enemy* does not appear[28] to be significantly influencing *Post-Reasoning* while *Cyber Event Awareness* exerts a positive and significant effect on this variable (∼0.44).

To validate the mediating role played by *Pre/Post-Reasoning*, mediation analysis is performed using the process suggested by Baron and Kenny [72]. For *Pre-Reasoning*, an Average Causal Mediation Effect (ACME) of 0.392 by *Enemy* is found to be statistically significant.[29] For *Post-Reasoning*, *Cyber Event Awareness* was found to fully mediate and has a significant[30] ACME of 0.815.

It should be noted that the new information presented in the preceding experiment provides participants with the possibility, but not certainty, that a non-state actor was responsible for the malicious behavior. Would participants behave differently if ambiguity is reduced? A second experiment is conducted with the difference being that ambiguity associated with the non-state actor is removed.

This follow-up experiment employs a sample of 152 participants[31] with a mean age of 39.8 years. In terms of *Risk Tolerance*, participants appear to be risk-averse given a mean of 0.326. With regards to *Cyber Event Awareness*, 0.735 indicates a fair awareness of events in cyberspace. Finally, a mean of 0.399 for *Exposure to Cyber Crime* indicates that participants have been victimized through cyberspace in one form or another. Samples for both experiments are comparable in terms of *Cyber Event Awareness* and *Exposure to Cyber Crime*, however; the later sample are found to be more risk averse. As with the first experiments, control variables were found to be balanced with the treatment.[32]

In terms of outcome variables, *Confidence* is comparable between the two experiments (7.40/7.20; $P = 0.287$). *Confidence Shift*, however, is statistically different with participants in the second experiment seemingly less confident of their previous assessment after being provided with new information ($-1.99/-2.79$; $P = 0.000$). With respect to the mediating variables, a similar pattern is observed. *Pre-Reasoning* values are statistically similar between both experiments (0.65/0.70 $P = 0.256$). *Post-Reasoning*, in contrast, shows that participants in Experiment 2 are more likely to question the responsibility to their perceived adversary[33] (0.48/0.39; $P = 0.012$). An explanation for this lies in the mediation analysis conducted on the *Post-Reasoning* variable. Unlike the first experiment, both *Enemy* and *Cyber Event Awareness* have a positive and significant influence on *Post-Reasoning*.[34] Mediation analysis further reveals that both these variables perfectly mediate *Post-Reasoning*.[35] Since neither of the above values are correlated ($P = 0.171$) one can assume that each of these exert their influence independently on *Post-Reasoning*.

The analysis reveals several key observations with respect to phenomenon of interest. First, images play a crucial role during the initial assessment and succeeding evaluation of information for decision-makers. Images, which represent specific beliefs, align judgements in situations where they are invoked. Furthermore, the presence of disconformity information does not completely mitigate the influence of beliefs. Second, images trigger a deliberative mechanism in the face of contradictory information. In such a situation, decision-makers engage in a deliberative process prior to the formation of judgement. Finally, this deliberative mechanism continues to manifest itself even in the absence of uncertainty. This suggests that decision-makers do not readily dismiss beliefs even in the face of certainty.

Previous behavior and cyberspace

Despite the persistent notion of cyber exceptionalism, judgements appear grounded in expectations established within the physical domain. Although its unique characteristics may not be ignored, the experimental results suggest a transference between expected behavior in the real world and that of cyberspace.

25 See Tables A4 and A5 in Appendix A.8 for complete details.
26 See Tables A2 and A3 in Appendix A.8 for complete details.
27 See Tables A6 and A7 in Appendix A.8 for complete details.
28 It is plausible to interpret this such that *Enemy* continues to mediate after the new information is presented but is much weaker compared to another variable that may not have been captured in the study. See Appendix A.10. In this case, $P = 0.07$ which suggests that if the effects are mild, a larger sample may result in a statistically significant result.
29 See Table A14 in Appendix A.10.
30 See Table A16 in Appendix A.10.
31 Participants in this follow-up experiment are also subjected to the same pre-analysis processing as the earlier group (e.g. attention checks and balancing).
32 See Table A1 in Appendix A.7.
33 The difference in this case points to a possible influence that the level of certainty has on moderating the impact of pre-existing beliefs. This is not specifically tackled in this article but is worth conducting at a later point in time.
34 See Table A13 in A.9.
35 See Tables A15 and A16 in Appendix A.10.

In their *Pre-Reasoning* response, participants in the treatment group drew parallels between past events and those described in the vignette. Observations such as "the hostilities attributed to the country are in line with the cyber-attack", "Vadare have a history of initiating attacks", and "given the trace and Vadare's prior behavior, they are a likely suspect" contribute substantially to the body of arguments that implicate Vadare. More importantly, participants hold firm to this belief despite the prospect of the contrary being true.

When the culpability of a third-party is surfaced, some participants chose interpretations that reinforce their beliefs and are aligned with the preceding judgement. For instance, one argues that the attribution appeared to make sense given Vadare's history. After reading the new information, the participant presented the possibility that a third-party may in fact be working with/for Vadare rather than acting independently. In other cases, participants simply ignored the new information and continued to re-assert their initial assessments. This point is crucial especially in the case of the first experiment. Despite the absence of a statistically significant differences between *Confidence Shift* and *Post-Reasoning*, continued references to past behavior, reframing, or dismissal of belief-incongruent information indicates a resistance to changing one's initial assessment.

Seizing on evidence that confirms pre-existing beliefs is unsurprising. As Dreyer [56] notes, constant exposure to multiple issues in a rivalry environment strengthens enemy images. More importantly, this exposure need not occur over the same issue. With cyberspace fast becoming an adjunctive foreign policy instrument, this reinforcement mechanism is likely to become more commonplace. This seizing effect, however, has several tangible, and negative, consequences if left unchecked.

The analysis of *Pre-Reasoning* arguments highlights an association between expected behavior and forensic evidence that support the attribution of the attack to Vadare. This is problematic given the inherent limitations of technical evidence surfaced by Lin [73]. Although artifacts such as IP addresses serve to identify the geographic location, questions of responsibility remain unanswered. Phrased differently, how can we be sure that an attack that originated from an adversary was indeed authorized by the current regime? Prior to the new information, some participants expressed the possibility that this may be a false flag operation meant to implicate Vadare.

During the operations against Estonia in 2007, officials were quick to assign blame to the Russian government for the initial wave of Denial-of-Service attacks. Considering the political atmosphere at the time and past Russian behavior, this is unsurprising. But the extent to which political elites held on to this idea despite the analysis provided by technical experts proving otherwise appears to confirm the findings presented herein [41].

While the act of misattribution is, by itself, harmless, policies resulting from such judgements are worrisome. The restraint observed among states in response to cyber operations may simply be a factor of the limited impact that these events demonstrate thus far. Even amid ongoing conflict in the physical domain, transgressions in cyberspace have yet to result in physical violence [74]. Yet with actors reserving the right to respond with kinetic options and coupled with the growing complexity and damage potential of these operations; unintended escalation remains a worrisome possibility. Moreover, with an expectation of increased interaction within this domain as reflected in the recent USA strategy and doctrine, misattribution could justify increased aggression that raises the risk of unintended escalation.

In a hypothetical case from Gartzke and Lindsay [9], a cyber operation aimed at an adversary's Nuclear Command, Control, and Communications (NC3) system may lead to conflict spirals should it be detected—and possibly mitigated. While this blunts the effects of the operation, this initiates a host of political and psychological factors on the side of the intended target. Specifically, the target may perceive an aggressor's hostile intent in response to this operation. Unfortunately, this perception may turn out to be incorrect given the malleability of cyber operations. As Buchanan [8] argues, tools that enable damaging cyber operations may also serve as instruments of espionage. Consequently, attempts to discern intent through the presence of such tools introduces uncertainty into the process. Furthermore, if one were to consider an aggressor's history, it would be easy to establish how a target could use pre-existing beliefs to overcome such uncertainty to establish intent. Depending on both strategic (international) and political (domestic) environments, the response to this event serves to complicate and aggravate matters.

With these in mind, the role of negative images in the process of *seizing* and *freezing* in response to cyber operations becomes clear. Short of an actual declaration of intent on the part of an aggressor, targets are uncertain of the motivations behind the event. In response, elites seize on accessible beliefs to curb uncertainty prior to formulating judgements. Once established, elites hold on to these beliefs even in the face of contradictory information to avoid cognitive dissonance and to reduce the consequences of reverting decisions that emerged from earlier judgements. Consequently, this validates both *Hypotheses 1 and 2*.

A little knowledge increases risk

To counteract these adverse effects, researchers assert that knowledge of the issue area serves to improve the quality of judgements [75]. Elites effectively employ cognitive shortcuts that complement their expertise [63, 75, 76]. Paradoxically, the results of this experiment suggest otherwise. Knowledgeable participants (i.e. greater familiarity with interstate interactions in cyberspace) appear less inclined to adjust their assessments in the face of new information. This does not contradict the existing literature in that the importance of knowledge is predicated on the alignment between currently held expertise and the real world. Given the opaque nature of cyber operations, the quality of the information pertaining to these events are questionable. Consequently, whatever flaws in judgement that emerge is not due solely to the need to maintain beliefs but is subject to the quality of information available. Expertise serve to improve the quality of judgement insofar as an individual's understanding of a phenomenon is correct [75, 77].

Expertise is a perennial problem in the attempt to better understand cyberspace as an instrument of foreign policy. At a fundamental level, basic knowledge of the underlying technology is limited [10, 78]. This results in several incorrectly held beliefs concerning the consequences of cyber operations, two of the most prominent being the low barriers of entry into the domain and the strategic efficacy of cyber operations [37].

Proponents of the revolutionary nature of cyberspace highlight its minimal material requirements. Citing the ease with which offensive tools may be obtained and the growing dependence on cyberspace, they argue that materially disadvantaged states are granted the means with which to negate these imbalances [12]. Yet, as convincing as this may seem, these claims remain unsubstantiated.

While cyber operations provide weaker states with additional foreign policy instruments, its utility is a function of its ability to

threaten cyber assets. Capable adversaries would, in theory, have in place defenses to contain threats from less technologically advanced adversaries. This effectively limits the strategic impact obtained through cyber operations and is readily observed through past incidents. Cyber operations capable of threatening an adversary's strategic interest are often associated with established powers due to the technological and organizational requirements needed to execute these operations effectively [7, 14].

Over the past decade, cyber operations consistently failed to meet their intended strategic objectives. As early as 2013, Iasiello [79] observed the limited gains achieved by employing cyber operations to shape foreign policy. Five years on, other researchers acknowledge these limits citing that <5% of operations intended to alter an adversary's behavior have succeeded [80]. Apart from the previously mentioned technological and organizational requirements, the constrained and transient effects of these operations are a major impediment to their strategic utility. Yet, despite the available evidence, the belief in the revolutionary potential of cyberspace is still widely held.

A study of news articles from 2008 to 2013 highlights several noteworthy observations regarding how the phenomenon is framed. Most notable is the persistent belief in the ability of cyber operations to inflict significant and lasting damage to infrastructure. The narrative for a potential "Cyber Armageddon" brought about by a malicious actor is still very much alive despite the absence of supporting evidence [24]. This encourages mistaken beliefs regarding the nature of cyber operations and its effects and result in sub-optimal judgements that may adversely affect elite decision-making.

Between certainty and ambiguity

A consistent observation surfaced through the experiments is the varying significance of *Enemy* and *Cyber Event Awareness* before and after new information is presented. In the case of the former, *Enemy* is the only variable that accounts for *Pre-Reasoning*. Whereas in the latter, the significance of *Enemy* on *Post-Reasoning* fluctuate between the two experiments. At a theoretical level, this is not particularly troublesome given the nature of beliefs and the use of schemas, a related cognitive structure.

Beliefs are most commonly represented through operational codes that are a set of beliefs about the political world and includes a conceptualization of the nature of politics and the means with which one's goals may be achieved [81]. By this definition, the behavior of a third party is perceived through these operational codes. Similarly, schemas are "cognitive structures that represents knowledge about a concept or type of stimulus, including its attributes and relations among those attributes" [82]. Both beliefs and schemas are cognitive structures that simplify information and whose quality varies with experience. The former, however, is limited to perceiving information, whereas the latter provides a deeper explanation of how information is retrieved, stored, and processed [48]. While it appears that these are two distinct constructs, they share similar processes to the extent that beliefs do not exist independent of schemas but rather that beliefs are encapsulated within schemas.

With respect to this article, three important features of schemas are crucial in understanding participant behavior. (i) First, when information is ambiguous, individuals form judgements using whichever schema is accessible [48]. (ii) Second, when information is missing, individuals use default values of the schema to fill in the existing gaps [83, 84]. (iii) Third, individuals will not interpret information to support a schema when the evidence that contradicts the schema is unambiguous [85].

Prior to being presented with new information, participants exposed to the treatment are made aware of Vadare's belligerent reputation. This, along with seemingly malicious actions of Vadare during the present dispute triggers this "enemy image" schema over other, more general schemas. This would account for the exclusive influence of *Enemy* on *Pre-Reasoning* observed in both experiments as this is possibly the most readily accessible, and unchallenged, schema given the absence of additional information. However, once the new information is presented, the differing levels of uncertainty prompts variation in the search for an appropriate schema. For the first experiment, mentioning the possible role of a non-state actor could trigger the search for a schema explaining how non-state actors behave under these conditions. This is directly associated with one's familiarity with these types of incidents and relate directly to *Cyber Event Awareness*. In the second experiment, however, emphasizing a degree of certainty that it was a non-state actor instead of Vadare may have provoked dissonance for participants in the treatment group. On the one hand the event is characteristic of non-state actors based on their knowledge. On the other hand, the current strategic environment points to Vadare as the party most likely to benefit. In this case, the varying effect sizes between *Enemy* (0.407) and *Cyber Event Awareness* (0.793) points to two competing schemas under consideration.[36] This is further reinforced in the actual text justifying their choice where these contrasting claims are laid out by participants.

Assuming that the most accessible schema is selected, participants begin an evaluative process resulting in specific judgements. This process does not necessarily require all information to be made available as schemas permit individuals to go "beyond" the information and to draw upon "expected" attributes based on what they have learned or experienced in the past [83]. Recurring references in both experiments to the unreliability of IP Addresses as a source of attribution, collusion between government and private actors, as well as the possibility of false flag operations are significant as none of these concepts are mentioned within the vignette. Consequently, it is fair to assume that participants include these arguments in their reasoning to address missing information to fully utilize their chosen schema. Furthermore, this utilization of information beyond what is provided accounts for the outcome of the second experiment despite the reduction in ambiguity.

Within interstate relations, evidence that completely contradicts pre-conceived notions are rare. Despite its wording, most participants in the second experiment did not consider the new information as irrefutable based on the open-ended responses provided. Instead, continued references are made for uncertainty and the continued possibility that this incident originated from Vadare. This observation aligns with theoretical expectations and the need to avoid cognitive dissonance and finds statistical support in the experimental results. Attempts to maintain the belief in Vadare's culpability is reflected more strongly in the second experiment than the first given

36 A statistically significant Average Direct Effect on *Post-Reasoning* was identified in the first experiment with respect to *Enemy*. This not only suggests that *Enemy* directly influences judgement without going through the mediating variable, but also hints at the possibly that another, as of yet unidentified variable, can account for *Post-Reasoning*.

the differences with respect to *Confidence Shift* ($P = 0.000$) and *Post-Reasoning* ($P = 0.014$) between these experiments as well as the mediation analysis on *Post-Reasoning*. Consequently, this confirms the assertions made by *Hypothesis 2*.

These preceding processes, while theoretically significant in our understanding of judgements vis-à-vis cyber operations, carry even greater consequences in the realm of policy. To begin with, suitable judgements emerge using an appropriate schema. In this case, appropriateness not only refers to the use of the correct/matching construct but that its underlying relations match reality. Schemas emerge either through direct or secondhand experience [48]. As such, accuracy is a function of this formative process.

As mentioned in the preceding subsection, the continued lack of expertise in the domain coupled with readily available information that may not faithfully reflect reality results in faulty knowledge of cyberspace. This is problematic on two fronts. First, an improper schema may result in false associations/relationships between available information. In the case of the vignette, limited information regarding Vadare's propensity for using cyber operations leads to an association between it and future incidents. In the real world, constant exposure to Russian and Chinese cyber operations may result in a narrow view of states active in this space. This raises the risk of false flag operations. Second, given that schemas allow individuals to go beyond the information provided; an ill-fitting schema propagates incorrect notions regarding the nature of cyber operations.

Conclusions and policy implications

The past decade has seen a growing interest in both the use and study of cyber operations. While cybersecurity scholarship continues to advance our understanding of this phenomenon, current research focuses primarily on technological and systemic factors to account for state behavior. Although crucial, these fail to note the importance of cognitive attributes that influence judgements that, in turn, result in policies that impact both state behavior and technological development. Thusly, this article serves to highlight the importance of the individual in interstate cyber interactions.

Through survey experiments, images are demonstrated to play a crucial role in the formation of judgements. Given that interactions between states within cyberspace do not occur within a strategic vacuum, past behavior serves as a complementary instrument used to reduce the uncertainty. Specifically, individuals rely on past adversarial behavior when evaluating incidents in cyberspace. While the use of such a heuristic serves as a crucial cognitive instrument, inappropriate usage results in poor policy responses.

The tendency of seizing on a pre-existing belief increases the susceptibility of a decision-maker to false-flag operations that, in turn, increases the likelihood of unintended escalation. In addition, confidence in the correctness of one's knowledge further increases the possibility of false-flag operations succeeding by limiting analysis to focus on technical evidence that can easily be falsified. Furthermore, freezing may aggravate the situation by encouraging decision-makers to form judgements that support existing beliefs but are not substantiated by evidence. Although, there have not been cases in which conflict in cyberspace spills into the physical domain, these effects are likely to increase tension between adversaries and could possibly extend the duration of the dispute in question.

As a clarificatory point, it crucial to note that degradative cyber operations are employed in these experiments as a means of studying behavior in the context of a worst-case scenario. There is no reason to dismiss the possibility that similar cognitive mechanisms responsible for *seizing* and *freezing* may also be manifested in cases of less damaging operations (i.e. disruptive or espionage). It is worth noting that given the frequency of these events, increased exposure may result in a moderation of biased judgement due to a greater familiarity with these incidents and the actors involved. However, barring research that explicitly investigate how consistent exposure to cyber operations influence attributive judgements, we can only speculate on the applicability of the above mechanisms on non-degradative operations.

While the article confirms existing theoretical arguments carried over into the cyber realm, it does provide two crucial observations with corresponding policy implications: analyzing events through the lens of history and the possibility of incorrect/inappropriate knowledge. With past-behavior serving as a clarificatory mechanism that minimizes the effects of uncertainty, policy makers should proceed with caution when attributing events—particularly those that would merit a response that carries significant political, strategic, economic, and/or military consequences. This propensity to misattribute, while not completely aligned with the scenario in this article, was recently demonstrated when rumors emerged that the wide-spread power outages in Argentina, Paraguay, and Uruguay were somehow linked to the recent claims that the USA had compromised the Russian power grid [86]. Familiarity with our over-dependence on pre-existing beliefs can serve to temper knee-jerk reactions that could prove more troublesome than the incident which prompted these.

Aside from pre-conceived beliefs regarding a potential adversary's culpability, the level of expertise held by policy makers is crucial at the onset of cyber security incidents. The continued knowledge gap between technology and policy experts increases the risk of inaccurate information driving policy choices. As such, initiatives to either educate key decision-makers on the fundamentals required to understand cyber operations or the inclusion of technology experts in the decision-making cycle are possible options to better utilize expertise in the formulation of judgements and the development of an appropriate response. To this end, decision-making models such as that provided by Rid and Buchanan [87] are helpful.

The attribution of cyber operations is not as problematic as it was a decade ago. That being said, technological developments alone do not result in sound judgement when bias-inducing processes continue unnoticed. Knowledge of our own fallibility encourages thoughtful policies that, hopefully, increase stability within this emergent domain.

Acknowledgements

The author would like to extend thanks to the members of the Digital Issues Discussion Group (DIDG) whose insights and comments served to shape and strengthen the arguments in this article.

Appendix A: 1. Experiment vignette

[Seen by all]

Your country, Idimore, has invested significantly in the development of an e-Government platform to provide public services via the Internet. Activities such as the filing of income tax returns, social benefit claims and voting during national elections are now available via an online platform accessed using personal computers or mobile devices.

Two days ago, national elections had to revert back to paper-based ballots due to a suspected cyber-attack. The National Cyber Security Center (NCSC) traced the attack back to government computers operated by your neighboring country, Vadare.

Treatment
Both Idimore and Vadare have been involved in a series of disputes in the past that are often initiated by Vadare. Vadare is known to pursue its interests using all possible means at its disposal that include, but are not limited to, trade embargoes, military intrusions across the border, and cyber-attacks. Over the past months, the president of Vadare has expressed support for a far-right candidate aspiring for the presidency in your country's upcoming national elections and has gone so far as to question the legitimacy of the current administration.

[Seen by all]
At present, the cyber-attack is still ongoing and has expanded to include your country's air traffic control systems. Several airports have reported disruptions in their operations due to the non-responsiveness of crucial systems. As the director of the NCSC, you are tasked with providing the president with the necessary information required to formulate an appropriate response.

[Only experiment 1]
Further forensic analysis reveals that the tools and techniques used for the cyber-attack are also employed by a hacktivist group called the CyberChaosCrew. This group has been known to actively target government systems for fun over the last 12 months.

[Only experiment 2][37]
Further forensic analysis reveals that a hacktivist group called the CyberChaosCrew is responsible for the cyber-attack. This group has been known to actively target government systems for fun over the last 12 months.

2. Confidence and reason measurements

Confidence (Pre)
How confident are you that the cyber-attack came from Vadare?
[0–10]

Pre-Reasoning
In 3–5 sentences, please elaborate on your answer to the previous question. The explanation should highlight the reason(s) behind your assessment of Vadare's role in this cyber-attack. Please provide your answers in English.
[Open-Ended]

Confidence (post)
Given the new information provided, how confident are you that Vadare is the source of the cyber-attack? [0–10]

Post-Reasoning
In 3–5 sentences, please elaborate on your answer to the previous question. The explanation should highlight the reason(s) you decided to keep or change your earlier assessment of Vadare's role in this cyber-attack. Please provide your answers in English.
[Open-Ended]

3. Risk measurements

Risk question 1
Some people say you should be cautious about making major changes in life. Suppose these people are located at 1. Others say that you will never achieve much in life unless you act boldly. Suppose these people are located at 7. And others have views in between. Where would you place yourself on this scale? [1–7]

Risk question 2
Suppose you were betting on horses and were a big winner in the third or fourth race. Would you be more likely to continue playing or take your winnings?
[Definitely continue playing (1), Probably continue playing (2),
Not sure (3), Probably take my winnings (4), Definitely take my winnings (5)]

Risk Question 3
I would like to explore strange places.
[Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5)]

Risk Question 4
I prefer friends who are exciting and unpredictable.
[Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5)]

Risk Question 5
I like new and exciting experiences, even if I have to break the rules.
[Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5)]

Risk Question 6
I like to do frightening things.
[Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5)]

Risk Question 7
In general, how easy or difficult is it for you to accept taking risks?
[Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5)]

Computation
$Risk = \text{Scale}(\text{Question 1}, [1, 5]) + \text{Question 2} + \ldots \text{Question 7})/7$
$Risk = \text{Scale}(Risk, [0, 1])$

37 The version for Experiment 2 differs in that a third-party's responsibility is less ambiguous. This was means to test whether the same processes observed in Experiment 1 is manifested under conditions of reduced uncertainty. This is not meant to be interpreted as a new treatment.

Interpretation
[0–0.5): Risk Averse
[0.5]: Balanced
(0.5–1.0]: risk acceptant

4. Cyber event questionnaire

Question 1
The cyber-attack against Estonia in 2007 disrupted banking systems within the small European nation. [**true** (1), false (0)]

Question 2
In December 2015, the USA suffered power outages due to a Russian-led cyber-attack.
[True (0), **false** (1)]

Question 3
The cyber-attack during the recently concluded 2018 Winter Olympics in Pyeongchang was incorrectly attributed to North Korea.
[**True** (1), false (0)]

Question 4
A botnet is a collection of compromised devices (e.g. computers, routers, etc.) that may be used to launch cyber-attacks on a large scale.
[**True** (1), false (0)]

Question 5
Phishing is an attempt to exhaust the resources of a targeted computer in order to cause it to fail.
[True (0), **False** (1)]

Question 6
These days, cyber-attacks can be caused by individuals, criminal organizations, or states.
[**True** (1), False (0)]

Computation
Cyber Event Awareness = (Question 1 + Question 2 + … Question 6)/6

Interpretation
[0–0.5): Unknowledgeable
[0.5]: Guessing
(0.5–1.0]: Knowledgeable

5. Cyber crime questionnaire

Question 1
Have any of your devices ever been infected with malware (e.g. virus, trojans, spyware, etc.)?
[Yes (1), No (0)]

Question 2
Have any of your online accounts (e.g. Google, Facebook, Twitter, etc.) ever been hacked?
[Yes (1), No (0)]

Question 3
Did you ever experience having your personal information (e.g. Social Security Number, Credit Card Number) stolen?
[Yes (1), No (0)]

Computation
Exposure to Cyber Crime = (Question 1 + Question 2 + Question 3)/3

Interpretation
[0–0.5): Underexposed
[0.5]: Moderately Exposed
(0.5–1.0]: Overexposed

6. Response codes

Code	Description	Type
Motive (Absence)	References to an actor's lack of motive to engage in cyber operations. No strategic or tactical benefit to be gained	Disconfirmatory
Organization (Trust)	References to the target's lack of organizational capability or the trustworthiness of the evaluating agency. No technical capabilities to accurately assess incidents or an ulterior motive.	Disconfirmatory
Information	References to the lack of information concerning the incident. No information that confirms culpability.	Disconfirmatory
False Flag	References to the possibility of a third-party being culpable. The chance that a false flag operation may have taken place.	Disconfirmatory
Technology (Limited)	References to the susceptibility of technology to manipulation. The limited nature of technology does not allow for accurate attribution.	Disconfirmatory
Capabilities	References to the capability of an actor to engage in cyber operations. They have the technological or military means to achieve their desired effect.	Confirmatory
Environment	References to the strategic environment that makes the actions of an actor more likely. The expectation of certain events as a continuation of an on-going exchange.	Confirmatory
Technology (Definite)	References to the trustworthiness of technology. The fact that forensic evidence exists that ties an actor to an event.	Confirmatory
History	References to an actor's past behavior. The expectation that an actor will continue to behave in a consistent manner.	Confirmatory
Organization (Competence)	References to an organization's competence in evaluating evidence. The trust in the material capabilities and professionalism of an organization.	Confirmatory
Motive (Presence)	References to an actor's motive to engage in cyber operations. The operation has some strategic or tactical utility.	Confirmatory
Coding Notes:	• Each expressed idea that matches the above code represents a single instance of that code. • A single sentence may be associated with 1 or more codes depending on the ideas communicated.	

Computation

Pre/Post-Reason = Total(Confirmatory)/Total(Confirmatory) + Total(Disconfirmatory)

Interpretation

[0–0.5): Doubtful
[0.5]: Even-Handed
(0.5–1.0]: Certain

7. Covariates balance table

Table A1. Covariate balance table

	Enemy—Experiment 1			*Enemy—Experiment 2*		
	Control	Treat	*P*	Control	Treat	*P*
Age	38.171	38.598	0.808	39.895	39.724	0.926
Cyber events	0.742	0.754	0.689	0.715	0.754	0.171
Cyber crime	0.451	0.431	0.657	0.373	0.425	0.251
Risk	0.435	0.434	0.962	0.335	0.317	0.348
Male	0.415	0.390	0.752	0.382	0.382	1
Five eyes	0.988	1.000	0.320	0.934	0.961	0.471
N	82	82		76	76	

8. Experiment 1 regression tables

Table A2. Confidence average direct effects models (EXP 1)

	Dependent variable: Confidence			
	(1)	(2)	(3)	(4)
Enemy	**0.549 (0.260)***	0.157 (0.233)	0.177 (0.229)	0.186 (0.229)
Pre-reasoning		**2.227 (0.309)*****	**2.226 (0.304)*****	**2.163 (0.305)*****
Cyber events			−0.015 (0.579)	−0.141 (0.596)
Cyber crime			**0.941 (0.395)***	**1.094 (0.397)****
Risk tolerance			0.734 (0.688)	0.803 (0.696)
Age				**0.023 (0.010)***
Male				0.093 (0.238)
Five eyes				−0.056 (1.449)
Constant	7.159 (0.184)***	5.905 (0.237)***	5.172 (0.578)***	4.351 (1.648)**
R^2	0.027	0.264	0.301	0.324
Adjusted R^2	0.021	0.255	0.279	0.289
Residual Std. Error	1.663 (df = 162)	1.451 (df = 161)	1.427 (df = 158)	1.417 (df = 155)
F Statistic	4.666 (df = 62; 1)	28.87 (df = 161; 2)	13.59 (df = 158; 5)	9.295 (df = 155; 8)

Note: $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Table A3. Confidence shift average direct effects model (EXP 1)

	Dependent variable: Confidence shift			
	(1)	(2)	(3)	(4)
Enemy	−0.537 (0.301)	**−0.778 (0.272)****	**−0.800 (0.272)****	**−0.817 (0.272)****
Post-reasoning		**1.968 (0.309)*****	**2.067 (0.318)*****	**2.026 (0.319)*****
Cyber events			−0.432 (0.711)	−0.168 (0.738)
Cyber crime			−0.713 (0.478)	−0.852 (0.482)
Risk tolerance			−0.658 (0.838)	−0.683 (0.851)
Age				−0.019 (0.012)
Male				−0.206 (0.290)
Five eyes				1.552 (1.758)
Constant	−1.720 (0.213)***	−2.551 (0.231)***	−1.665 (0.670)**	−2.504 (1.984)
R^2	0.019	0.216	0.234	0.253
Adjusted R^2	0.013	0.206	0.210	0.214
Residual std. error	1.925 (df=162)	1.726 (df=161)	1.722 (df=158)	1.717 (df=155)
F Statistic	3.186 (df=162; 1)	22.23 (df=161; 2)	9.673 (df=158; 5)	6.561 (df=155; 8)

Note: $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Table A4. Confidence total effects models (EXP 1)

	Dependent variable: Confidence		
	(1)	(2)	(3)
Enemy	**0.549 (0.260)***	**0.568 (0.257)***	**0.575 (0.255)***
Cyber events		0.038 (0.667)	−0.132 (0.684)
Cyber crime		**0.902 (0.455)***	**1.098 (0.456)***
Risk tolerance		0.866 (0.793)	1.005 (0.797)
Age			**0.029 (0.012)***
Male			0.065 (0.274)
Five eyes			−0.981 (1.656)
Constant	7.159 (0.184)***	6.346 (0.640)***	6.143 (1.869)**
R^2	0.027	0.064	0.105
Adjusted R^2	0.029	0.040	0.064
Residual std. error	1.663 (df=162)	1.646 (df=159)	1.625 (df=156)
F Statistic	4.466 (df=162; 1)	2.72 (df=159; 4)	2.603 (df=156; 7)

Note: $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Table A5. Confidence shift total effects models (EXP 1)

	Dependent variable: Confidence Shift		
	(1)	(2)	(3)
Enemy	−0.537 (0.301)	−0.554 (0.302)	−0.593 (0.312)
Cyber events		0.441 (0.784)	0.794 (0.809)
Cyber crime		−0.596 (0.535)	−0.749 (0.539)
Risk tolerance		0.091 (0.932)	0.067 (0.943)
Age			−0.019 (0.014)
Male			−0.309 (0.324)
Five eyes			2.598 (1.959)
Constant	−1.720 (0.213)***	0.091 (0.932)*	−3.729 (2.210)
R^2	0.019	0.030	0.058
Adjusted R^2	0.013	0.005	0.016
Residual std. error	1.925 (df=162)	1.933 (df=159)	1.922 (df=156)
F Statistic	3.186 (df=162; 1)	1.211 (df=159; 4)	1.379 (df=156; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A6. Pre-reasoning mediation models (EXP 1)

	Dependent variable: Pre-reasoning		
	(1)	(2)	(3)
Enemy	**0.176 (0.058)****	**0.175 (0.058)****	**0.180 (0.058)****
Cyber events		0.024 (0.151)	0.004 (0.157)
Cyber crime		−0.018 (0.103)	0.002 (0.104)
Risk tolerance		0.059 (0.179)	0.093 (0.183)
Age			0.003 (0.003)
Male			−0.013 (0.063)
Five eyes			−0.428 (0.379)
Constant	0.563 (0.041)***	0.527 (0.145)***	0.828 (0.428)
R^2	0.055	0.056	0.072
Adjusted R^2	0.049	0.032	0.030
Residual std. error	0.369 (df=162)	0.372 (df=159)	0.372 (df=156)
F Statistic	9.354 (df=162; 1)	2.336 (df=159; 4)	1.727 (df=156; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A7. Post-reasoning mediation models (EXP 1)

	Dependent variable: Post-reasoning		
	(1)	(2)	(3)
Enemy	0.123 (0.069)	0.119 (0.067)	0.111 (0.068)
Cyber events		**0.422 (0.174)***	**0.475 (0.182)****
Cyber crime		0.057 (0.119)	0.051 (0.121)
Risk tolerance		0.362 (0.207)	0.370 (0.212)
Age			0.000 (0.003)
Male			0.051 (0.073)
Five eyes			0.516 (0.440)
Constant	0.423 (0.048)***	−0.074 (0.167)	−0.604 (0.496)
R^2	0.019	0.077	0.087
Adjusted R^2	0.013	0.053	0.046
Residual std. error	0.439 (df=162)	0.430 (df=159)	0.431 (df=156)
F Statistic	3.213 (df=162; 1)	3.294 (df=159; 4)	2.117 (df=156; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

9. Experiment 2 regression tables

Table A8. Confidence average direct effects models (EXP 2)

	Dependent variable: Confidence			
	(1)	(2)	(3)	(4)
Enemy	**0.592 (0.271)***	−0.003 (0.990)	−0.014 (0.200)	0.022 (0.199)
Pre-reasoning		**3.849 (0.317)*****	**3.919 (0.315)*****	**3.867 (0.314)*****
Cyber events			0.887 (0.544)	0.681 (0.555)
Cyber crime			**−0.731 (0.344)***	**−0.713 (0.342)***
Risk tolerance			−0.228 (0.828)	0.080 (0.848)
Age				**0.018 (0.009)***
Male				−0.042 (0.198)
Five eyes				−0.473 (0.432)
Constant	6.934 (0.192)***	4.556 (0.239)***	4.227 (0.540)***	4.029 (0.753)***
R^2	0.031	0.512	0.535	0.551
Adjusted R^2	0.024	0.506	0.519	0.526
Residual std. error	1.673 (df=150)	1.19 (df=149)	1.174 (df=146)	1.166 (df=143)
F Statistic	4.762 (df=150; 1)	78.24 (df=149; 2)	33.63 (df=146; 5)	21.93 (df=143; 8)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A9. Confidence shift average direct effects models (EXP 2)

	Dependent variable: Confidence shift			
	(1)	(2)	(3)	(4)
Enemy	−0.132 (0.335)	−0.538 (0.328)	−0.566 (0.332)	−0.595 (0.336)
Post-reasoning		**2.392 (0.534)*****	**2.532 (0.546)*****	**2.580 (0.553)*****
Cyber events			−1.052 (0.916)	−0.896 (0.944)
Cyber crime			0.585 (0.572)	0.579 (0.578)
Risk tolerance			−0.830 (1.371)	− 1.062 (1.420)
Age				−0.014 (0.015)
Male				−0.012 (0.332)
Five eyes				0.323 (0.729)
Constant	−2.724 (0.237)***	−3.427 (0.273)***	−2.655 (0.840)**	−2.434 (1.245)
R^2	0.001	0.120	0.134	0.140
Adjusted R^2	−0.006	0.108	0.105	0.092
Residual std. error	2.063 (df=150)	1.943 (df=149)	1.947 (df=146)	1.960 (df=143)
F Statistic	0.155 (df=150; 1)	10.14 (df=149; 2)	4.526 (df=146; 5)	2.916 (df=143; 8)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A10. Confidence total effects models (EXP 2)

	Dependent variable: Confidence		
	(1)	(2)	(3)
Enemy	**0.592 (0.271)***	**0.612 (0.276)***	**0.644 (0.275)***
Cyber events		0.484 (0.778)	0.245 (0.792)
Cyber crime		−0.560 (0.492)	−0.529 (0.489)
Risk tolerance		0.532 (1.182)	1.062 (1.207)
Age			**0.028 (0.013)***
Male			−0.164 (0.282)
Five eyes			−0.391 (0.617)
Constant	6.934 (0.192)***	6.619 (0.722)***	5.929 (1.054)***
R^2	0.031	0.042	0.076
Adjusted R^2	0.024	0.016	0.031
Residual std. error	1.673 (df=150)	1.68 (df=147)	1.667 (df=144)
F Statistic	4.762 (df=150; 1)	1.605 (df=147; 4)	1.683 (df=144; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A11. Confidence shift total effects model (EXP 2)

	Dependent variable: Confidence shift		
	(1)	(2)	(3)
Enemy	−0.132 (0.335)	−0.158 (0.342)	−0.166 (0.346)
Cyber events		−0.285 (0.962)	−0.209 (0.997)
Cyber crime		0.362 (0.609)	0.350 (0.615)
Risk tolerance		−1.037 (1.463)	−1.265 (1.518)
Age			−0.010 (0.016)
Male			0.060 (0.355)
Five eyes			−0.016 (0.776)
Constant	−2.724 (0.237)***	−2.307 (0.894)*	−1.885 (1.326)
R^2	0.001	0.007	0.010
Adjusted R^2	−0.006	−0.021	−0.038
Residual std. error	2.063 (df=150)	2.079 (df=147)	2.097 (df=144)
F Statistic	0.155 (df=150; 1)	0.242 (df=147; 4)	0.201 (df=144; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A12. Pre-reasoning mediation models (EXP 2)

	Dependent variable: Pre-reasoning		
	(1)	(2)	(3)
Enemy	**0.154 (0.050)****	**0.160 (0.051)****	**0.161 (0.051)****
Cyber events		−0.103 (0.142)	−0.113 (0.147)
Cyber crime		0.044 (0.090)	0.048 (0.091)
Risk tolerance		0.192 (0.216)	0.254 (0.224)
Age			0.002 (0.002)
Male			−0.031 (0.052)
Five eyes			0.021(0.114)
Constant	0.618 (0.035)***	0.610 (0.132)***	0.491 (0.195)**
R^2	0.061	0.071	0.081
Adjusted R^2	0.054	0.046	0.037
Residual std. error	0.306 (df=150)	0.308 (df=147)	0.309 (df=144)
F Statistic	9.669 (df=150; 1)	2.825 (df=147; 4)	1.825 (df=144; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

Table A13. Post-reasoning mediation models (EXP 2)

	Dependent variable: Post-reasoning		
	(1)	(2)	(3)
Enemy	**0.170 (0.048)*****	**0.161 (0.048)****	**0.166 (0.049)*****
Cyber events		**0.303 (0.136)***	0.266 (0.140)
Cyber crime		−0.088 (0.086)	−0.089 (0.087)
Risk tolerance		−0.082 (0.207)	−0.079 (0.214)
Age			0.002 (0.002)
Male			0.028 (0.050)
Five eyes			−0.131 (0.109)
Constant	0.294 (0.034)***	0.138 (0.127)	0.213 (0.187)
R^2	0.076	0.114	0.126
Adjusted R^2	0.070	0.090	0.084
Residual std. error	0.297 (df=150)	0.294 (df=147)	0.295 (df=144)
F Statistic	12.42 (df=150; 1)	4.711 (df=147; 4)	2.977 (df=144; 7)

Note: $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

10. Mediation tables

Table A14. Enemy > post-reasoning > confidence mediation effects

	Enemy > Pre-Reasoning > Confidence							
	Experiment 1				Experiment 2			
	Estimate	CI Lower	CI Higher	*p*	Estimate	CI Lower	CI Higher	*p*
ACME	**0.392**	**0.131**	**0.660**	**0.004**	**0.595**	**0.242**	**1.010**	**0.000**
ADE	0.157	−0.318	0.600	0.444	−0.003	−0.384	0.360	0.948
Total effects	0.549	−0.005	1.040	0.060	**0.592**	**0.078**	**1.130**	**0.028**
Pr. mediated	0.714	−0.904	2.990	0.064	**1.004**	**0.521**	**2.760**	**0.028**

Table A15. Enemy > post-reasoning > confidence shift mediation effects

	Enemy > Post-Reasoning > Confidence Shift							
	Experiment 1				Experiment 2			
	Estimate	CI Lower	CI Higher	*P*	Estimate	CI Lower	CI Higher	*p*
ACME	0.242	−0.009	0.510	0.072	**0.407**	**0.141**	**0.730**	**0.000**
ADE	**−0.778**	**−1.377**	**−0.210**	**0.000**	−0.538	−1.163	0.070	0.072
Total Effects	−0.537	−1.184	0.100	0.092	−0.132	−0.777	0.510	0.616
Pr. Mediated	−0.450	−3.634	2.480	0.164	−3.090	−23.461	12.910	0.616

Table A16. Cyber events > post-reasoning > confidence shift mediation effects

	Cyber Event > Post-Reasoning > Confidence Shift							
	Experiment 1				Experiment 2			
	Estimate	CI Lower	CI Higher	*P*	Estimate	CI Lower	CI Higher	*p*
ACME	**0.815**	**0.179**	**1.54**	**0.020**	**0.793**	**0.201**	**1.630**	**0.004**
ADE	−0.353	−1.477	0.770	0.500	−1.075	−2.763	0.850	0.228
Total Effects	0.462	−0.765	1.770	0.440	−0.282	−2.200	1.630	0.748
Pr. Mediated	1.764	−13.274	10.830	0.440	−2.813	−9.476	14.700	0.752

References

1. Gray CS. *Making Strategic Sense of Cyber Power: Why the Sky is not Falling*. Pennsylvania: Strategic Studies Institute, 2013.
2. Kruglanski AW, Gigerenzer G. Intuitive and deliberate judgments are based on common principles. *Psychol Rev* 2011;**118**:97–109.
3. Jervis R. *Perception and Misperception in International Politics*. Princeton: Princeton University Press, 1976.
4. Chong D. Degree of rationality in politics. In: Huddy Ls (ed.), *The Oxford Handbook of Political Psychology*. Oxford: Oxford University Press, 2013, 96–129.
5. Rousseau DL, Garcia-Retamero R. Identity, power, and threat perception - a cross-national experimental study. *J Confl Resol* 2007;**51**:744–71.
6. Herrmann RK, Voss JF, Schooler TYE *et al*. Images in international relations: an experimental test of cognitive schemata. *Int Stud Quart* 1997;**41**: 403–33.
7. Valeriano B, Maness RC. *Cyber War versus Cyber Realities: Cyber Conflict in the International System*. Oxford; New York: Oxford University Press, 2015.
8. Buchanan B. *The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations*. London: Hurst & Company, 2017.
9. Gartzke E, Lindsay JR. Thermonuclear cyberwar. *J Cybersecurity* 2017;**3**: 37–48.
10. Dean B, McDermott R. A research agenda to improve decision making in cyber security policy. *Penn State J Law Int Affairs* 2017;**5**:29–164.
11. Valeriano B, Jensen B, Rc M. *Cyber Strategy: The Evolving Character of Power and Coercion*. New York: Oxford University Press, 2018.
12. Borghard ED, Lonergan SW. *The Logic of Coercion in Cyberspace*. *Security Stud* 2017;**26**:452–81.
13. Lindsay J, Gartzke E. Coercion through cyberspace: the stability-instability paradox revisited. In: Greenhill K, Krause Ps (eds), *The Power to Hurt: Coercion in the Modern World*. New York: Oxford University Press, 2014.
14. Slayton R. What is the cyber offense-defense balance? Conceptions, causes, and assessment. *Int Security* 2017;**41**:72–109.
15. Libicki MC. *Cyberdeterrence and Cyberwar*. Santa Monica: Rand Corporation, 2009.
16. Perrow C. *Normal Accidents: Living with High Risk Technologies*. Princeton University Press, 1984.
17. Dunn Cavelty M. From cyber-bombs to political fallout: threat representations with an impact in the cyber-security discourse. *Int Stud Rev* 2013; **15**:105–22.
18. Saltzman I. Cyber posturing and the offense-defense balance. *Contemp Security Pol* 2013;**34**:40–63.
19. Gomez MA, Villar EB. Fear, uncertainty, and dread: cognitive heuristics and cyber threats. *Polit Govern* 2018;**6**:61–72.

20. Camerer CF, Kunreuther H. Decision-processes for low probability events - policy implications. *J Pol Anal Manage* 1989;**8**:565–92.
21. Gigerenzer G. Out of the frying pan into the fire: behavioral reactions to terrorist attacks. *Risk Anal* 2006;**26**:347–351.
22. Reinhardt GY. Imagining worse than reality: comparing beliefs and intentions between disaster evacuees and survey respondents. *J Risk Res* 2017; **20**:169–94.
23. Valeriano B, Maness RC. The dynamics of cyber conflict between rival antagonists, 2001-11. *J Peace Res* 2014;**51**:347–60.
24. Jarvis L, Macdonald S, Whiting A. Unpacking cyberterrorism discourse: specificity, status, and scale in news media constructions of threat. *Eur J Int Security* 2017;**2**:64–87.
25. Huddy LS, David O, Levy J. Introduction: political psychology fundamentals. In: Huddy LS (ed.), *The Oxford Handbook of Political Psychology*. Oxford: Oxford University Press, 2013, xvii, 986 pages.
26. Mintz A, DeRouen, K. Jr. *Psychological Factors Affecting Foreign Policy Decisions. Understanding Foreign Policy Decision Making*. New York: Cambridge University Press, 2010, 114–18.
27. Stein JG. The micro-foundations of international relations theory: psychology and behavioral economics. *Int Organ* 2017;**71**:S249–63.
28. Goldstein DG, Gigerenzer G. The recognition heuristic how ignorance makes us smart. In: Gigerenzer G, Todd PM, Group TARs (eds), *Simple Heuristics That Make us Smart*. New York: Oxford University Press, 1999, 37–58.
29. Kahneman D, Slovic P, Tversky A. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge; New York: Cambridge University Press, 1982.
30. McDermott R. The psychological ideas of Amos Tversky and their relevance for political science. *J Theor Pol* 2001;**13**:5–33.
31. Mercer J. Rationality and psychology in international politics. *Int Organ* 2005;**59**:77–106.
32. Simon HA. Invariants of human behavior. *Ann Rev Psychol* 1990;**41**: 1–20.
33. Gigerenzer G. Why heuristics work. *Persp Psychol Sci* 2008;**3**:20–29.
34. Todd P, Gigerenzer G. *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press, 2012.
35. Jervis R. Understanding beliefs and threat inflation. *American Foreign Policy and the Politics of Fear: Threat Inflation since* 2009; **9**:16–39.
36. Van Evera S. Offense, defense, and the causes of war. *Int Security* 1998; **22**:5–43.
37. Liff AP. Cyberwar: a new 'absolute weapon'? The proliferation of cyberwarfare capabilities and interstate war. *J Strat Stud* 2012;**35**:401–428.
38. Axelrod R, Iliev R. Timing of cyber conflict. *Proc Natl Acad Sci* 2014; **111**:1298–303.
39. Hare F. The cyber threat to national security: why can't we agree?. *Conference on Cyber Conflict, Proceedings 2010* 2010; 211–25.
40. Rivera J. Achieving Cyberdeterrence and the Ability of Small States to Hold Large States at Risk. *2015 7th International Conference on Cyber Conflict - Architectures in Cyberspace (Cycon)* 2015:7–24.
41. Hansel M. Cyber-attacks and psychological IR perspectives: explaining misperceptions and escalation risks. *J Int Relat Dev* 2016;**21**:1–29.
42. Rascagneres P, Lee M. *Who Wasn't Responsible for Olympic Destroyer? Talos Intelligence*: Talos, 2018.
43. Power R. *The Solar Sunrise Case: Mak, Stimpy, and Analyzer Give the DoD a Run for Its Money.* http://www.informit.com/articles/article.aspx?p=19603&seqNum=4 (11 September 2019, date last accessed).
44. Holmes M. Believing this and alieving that: theorizing affect and intuitions in international politics. *Int Stud Quart* 2015;**59**:706–20.
45. Mercer J. Emotional beliefs. *Int Organ* 2010;**64**:1–31.
46. Roach SC. Affective values in international relations: theorizing emotional actions and the value of resilience. *Politics* 2016; **36**:400–12.
47. Sasley BE. Affective attachments and foreign policy: Israel and the 1993 Oslo Accords. *Eur J Int Relat* 2010;**16**:687–709.
48. Larson DW. The role of belief systems and schemas in foreign policy decision-making. *Pol Psychol* 1994;**15**:17–33.
49. Lane RE. *Political Ideology: Why the American Common Man Believes What he Does* 1962. New York: The Free Press.
50. Knutson JN. *The Human Basis of the Polity: A Psychological Study of Political Men*. Chicago: Aldine-Atherton, 1972.
51. Lodge M, Taber C. Three steps toward a theory of motivated political reasoning. In: Lupia A, McCubbins MD, Popkin S, L.s (eds), *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. Cambridge, UK: Cambridge University Press, 2000, 183–213.
52. Taber CS, Lodge M, Glathar J. The motivated construction of political judgments. In: Kuklinski JHs (ed.), *Citizens and Politics: Perspectives from Political Psychology*. Cambridge: Cambridge University Press, 2001, 198–226.
53. Boulding KE. National images and international systems. *J Confl Resol* 1959;**3**:120–31.
54. Holsti OR. The belief system and national images: a case study. *J Confl Resol* 1962; **6**:244–52.
55. Holsti OR. Cognitive dynamics and images of the enemy. *J Int Affairs* 1967;**21**:16–39.
56. Dreyer DR. Issue conflict accumulation and the dynamics of strategic rivalry. *Int Stud Quart* 2010;**54**:779–795.
57. Blum SC, Silver RC, Poulin MJ. Perceiving risk in a dangerous world: associations between life experiences and risk perceptions. *Soc Cogn* 2014;**32**:297–314.
58. Kruglanski AW, Webster DM. Motivated closing of the mind: "Seizing" and" freezing". *Psychol Rev* 1996;**103**:263.
59. Lerner JS, Tetlock PE. Accounting for the effects of accountability. *Psychol Bull* 1999;**125**:255–75.
60. Bar-Joseph U, Kruglanski AW. Intelligence failure and need for cognitive closure: on the psychology of the Yom Kippur surprise. *Pol Psychol* 2003; **24**:75–99.
61. Gomez MA. Sound the alarm! Updating beliefs and degradative cyber operations. *Eur J Int Security* 2019;**4**:1–19.
62. Kostyuk N, Wayne C, Communicating C. *Yber-Security: Citizen Risk Perception of Cyber-Threats*. Working Paper, 2018.
63. Hafner-Burton EM, Hughes AD, Victor DG. The cognitive revolution and the political psychology of elite decision making. *Persp Pol* 2013;**11**: 368–86.
64. Peer E, Brandimarte L, Samat S *et al*. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 2017;**70**: 153–63.
65. Crump M, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One* 2013;**8**:e57410.
66. Lupia A. New ideas in experimental political science. *Pol Anal* 2002;**10**: 319–24.
67. Mintz A, Redd SB, Vedlitz A. Can we generalize from student experiments to the real world in political science, military affairs, and international relations?. *J Confl Resol* 2006;**50**:757–76.
68. Galinsky AD, Gruenfeld DH, Magee JC. From power to action. *J Personal Soc Psychol* 2003;**85**:453.
69. Galinsky AD, Magee JC, Inesi ME *et al*. Power and perspectives not taken. *Psychol Sci* 2006;**17**:1068–74.
70. Inesi EM. Power and loss aversion. *Organiz Behav Human Decis Proc* 2010;**112**:58–69.
71. Kam CD, Simas EN. Risk orientations and policy frames. *J Pol* 2010;**72**: 381–96.
72. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personal Soc Psychol* 1986;**51**:1173.
73. Lin HS. Attribution of malicious cyber incidents: from soup to nuts. *J Int. Aff*. 2016;**70**:75–137.
74. Kostyuk N, Zhukov YM. Invisible digital front: can cyber attacks shape battlefield events?. *J Confl Resol* 2017;**45**:951–71.
75. Lau R, Redlawsk DP. Advantages and disadvantages of cognitive heuristics in political decision making. *Am J Pol Sci* 2001;**45**:951–71.
76. Saunders EN. No substitute for experience: presidents, advisers, and information in group decision making. *Int Organ* 2017;**71**:S219–47.
77. Kruglanski AW, Orehek E, Dechesne M *et al*. Lay epistemic theory: the motivational, cognitive, and social aspects of knowledge formation. *Soc Personal Psychol Compass* 2010;**4**:939–50.

78. Hansen L, Nissenbaum HD. Disaster, cyber security, and the Copenhagen School. *Int Stud Quart* 2009;**53**:1155–75.
79. Iasiello E. Cyber attack: a dull tool to shape foreign policy. In: Podins K, Stinissen J, Maybaum MS (eds), *2013 5th International Conference on Cyber Conflict*. Tallinn: IEEE, 2013, 451–70.
80. Jensen B, Maness RC, Valeriano B. Cyber Victory: the Efficacy of Cyber Coercion. *Annual Meeting of the International Studies Association* 2016. Atlanta: International Studies Association.
81. George AL. The "operational code": a neglected approach to the study of political leaders and decision-making. *Int Stud Quart* 1969;**13**:190–222.
82. Crocker J, Fiske ST, Taylor SE. *Schematic Bases of Belief Change. Attitudinal Judgment*. New York: Springer, 1984, 197–226.
83. Markus H, Zajonc RB. The cognitive perspective in social psychology. *Handbook of Social Psychology* 1985; **1**:137–230.
84. Ortony A, Rumelhart DE. *The Representation of Knowledge in Memory. Schooling and the Acquisition of Knowledge: Routledge*, 2017, 99–135.
85. Fiske ST, Taylor SE. *Social Cognition: From Brains to Culture*. London: Sage, 2013.
86. Fickling D. *Cyberattacks Make Smart Grids Look Pretty Dumb*. https://www.bloomberg.com/opinion/articles/2019-06-17/argentina-blaming-hackers-for-outage-makes-smart-grids-look-dumb (date last accessed)
87. Rid T, Buchanan B. Attributing cyber attacks. *J Strat Stud* 2015;**38**: 4–37.