New Perspectives
2021, Vol. 29(2) 144–164

DOI: 10.1177/2336825X21995702
journals.sagepub.com/home/nps

Grey matters: Advancing a psychological effects-based approach to countering malign information influence

Aiden Hoyle
University of Amsterdam, The Netherlands; TNO (Netherlands Organization for Applied Scientific Research), The Netherlands; Netherlands Defense Academy, The Netherlands

Helma van den Berg
TNO (Netherlands Organization for Applied Scientific Research), The Netherlands

Bertjan Doosje
University of Amsterdam, The Netherlands

Martijn Kitzen
Netherlands Defense Academy, The Netherlands

Abstract

Hostile political actors frequently engage in malign information influence, projecting antagonistic strategic narratives in targeted societies to manipulate the information environment and distort the perceptions of the citizens. Research examining malign information influence is growing, but more attention could be given to its psychological effects. Information operations are commonly assumed to affect the levels of trust and the emotional experiences of citizens who are targeted by them, but these notions are currently supported by limited evidence. We propose that experimental psychological research is a promising avenue to more clearly demonstrate these effects and individual differences of the target audience that may exacerbate these effects. This article discusses the knowledge gap regarding the psychological effects of malign information influence and suggests relevant psychological research that can be built upon when devising experimental studies that might address it. Finally, the article outlines key benefits that insights gleaned from this experimental research would offer to those seeking to counter malign information influence.

Corresponding author:
Aiden Hoyle, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WS Amsterdam, The Netherlands.
Email: a.hoyle@uva.nl

Keywords
Grey zone, hybrid warfare, information influence, information warfare, psychology, strategic narratives

Introduction

Information operations are increasingly being used by state and non-state political actors as a complementary or alternative means to traditional, kinetic warfare tactics (Ducheine, 2014: 3). 'Malign information influence' denotes this operational use of 'information as a weapon to inflict harm upon others by authoritarian regimes or other hostile actors' (Wagnsson, 2020: 1). In this, it captures hostile actors' use of targeted operations to promote crafted messaging that is designed to influence citizens in foreign states in a strategically desirable manner (Wilson et al., 2018). Information operations are not novel; print, television, radio media, and cultural institutions have been used by political actors to influence targeted societies since the Cold War era (Webb, 2014). Yet technological developments have created an additional suite of methods, such as online news outlets or social media platforms, through which hostile actors can effectively conduct malign information influence, unrestrained by borders, language, or resource limitations (Bachmann and Gunneriusson, 2015: 84). Such influence activities enable actors to pursue political and strategic goals without necessitating the use of traditional military force or the breaching of formalised aggression thresholds. Consequently, an amorphous security landscape is created without clear definitions of war and peace or well-defined demarcations of battlefields. This notion is a key aspect of emerging – yet contested – concepts of contemporary warfare such as so-called 'information' warfare, 'hybrid' warfare, or 'grey-zone' warfare (Almäng, 2019; Bressan and Sulg, 2020).

The explosion in potential avenues through which hostile actors can conduct malign information influence has been mirrored by both mounting concerns regarding the security challenges it poses to societies and research seeking to address this challenge. This has included research efforts to understand how social media is used to disseminate hostile narratives (e.g. Starbird et al., 2019) and what these overarching narratives are (e.g. Rebegea, 2019), to investigations into the psychological mechanisms enabling disinformation belief (e.g. Lewandowsky et al., 2012). Yet an aspect that has, so far, attracted less attention is that of the psychological effects that malign information influence has in targeted citizens. In this article, we argue that evidence of the psychological effects of malign information influence is critical to gaining a broader understanding of malign information influence and that this understanding would offer useful practical insights for different counter efforts.

We begin the article by discussing the concept of malign information influence, how antagonistic strategic narratives are integral to such information influence activities, and relevant examples of such activities. We follow by outlining psychology's current contribution to our understanding of malign information influence, before underlining how the lack of evidence regarding the psychological effects of malign information influence constitutes a crucial knowledge gap. Next, we sketch the borders of a research agenda that centres experimental psychological research as a possible avenue to address this gap. Here, we move through the empirical process, providing suggestions for relevant psychological factors, the generation of theoretical models and experimental designs. In this section, we parse key studies regarding conceptually similar phenomena, such as conspiracy theories or populist media narratives, that can be used as a foundation to devise this experimental research. We conclude the article by suggesting four practical benefits that this experimental research can offer to those seeking to counter malign information influence.

Malign information influence and antagonistic strategic narratives

A primary goal of malign information influence is to steer perceptions and discourse regarding the political reality in a direction that facilitates the pursuit of the hostile actors' strategic or political goals (Wilson et al., 2018: 2). To do so, information operations centre around pushing what has been described as 'antagonistic' strategic narratives (Wagnsson and Barzanje, 2019: 2). Strategic narratives are defined as 'a means for political actors to construct a shared meaning of the past, present, and future of international relations in order to shape the opinions and behaviour of actors at home and overseas' (Miskimmon et al., 2014: 176). They are understood as 'compelling storylines which can explain events convincingly and from which inferences can be drawn' (Freedman, 2006: 22), and can be constructed by varying levels of factual information, information that is misleading, oversimplified, or provocative, and information that is intentionally false – the latter being termed as 'disinformation' (Betz, 2008: 514; Flore, 2020: 21; Wilson et al., 2018: 3). Strategic narratives are thought to operate on many levels (Betz, 2008: 519), of which one can be assumed to be psychological. Narratives are considered as cognitive lenses through which people view, interpret, and make meaning of the social world (Lewandowsky et al., 2013: 488). With the reality of international (geo)politics being complex and intricate, narratives are key to imbuing regular citizens' perceptions of this reality with structure and coherence (Antoniades et al., 2010: 5). Strategic narratives are therefore seen as 'power resources' for political actors, who can tactically influence large populations by shaping how citizens interpret information about the (international) political order (Roselle et al., 2014: 74).

Strategic narratives can be used by political actors for a range of different objectives (Colley, 2017: 1). Often, this involves the control of their own image in the international arena by promoting specific narratives that are favourable for them (Antoniades et al., 2010: 5). Recent examples of this include Russian information operations that promoted narratives obfuscating their role in the MH17 disaster (Rietjens, 2019) or the Skripal incident (Ramsay and Robertshaw, 2019), and Chinese information operations disseminating narratives discrediting the Hong Kongese protestors (la Cour, 2020). However, strategic narratives can also be used to achieve even more malign objectives of different political actors, with information operations promoting what have been described as 'hostile' or 'antagonistic' strategic narratives (Deverell et al., 2020: 2; Flore, 2020: 2; Wagnsson and Barzanje, 2019: 2). Using antagonistic strategic narratives, hostile actors can weaponise public opinion and coerce decision-making in foreign societies by confounding discourse in a way that destabilises their civilian domain (Pomerantsev and Weiss, 2014: 4; Wilson et al., 2018: 2). As such, antagonistic strategic narratives typically exploit vulnerabilities in a foreign society, such as inequalities, latent historic grievances, contentious societal issues, or spontaneous moments of societal uncertainty or tension (Flore, 2020: 13; Lanoszka, 2016: 182–186).

Russian information operations that targeted latent grievances in the Russian communities of Ukrainian regions of Donetsk, Luhansk, and Crimea, prior to the 2014 crisis, provide a prime example of this strategy (Ambrosio, 2016; Khaldarova and Pantti, 2016). Russia is thought to have weaponised 'information, culture and money' (Pomerantsev and Weiss, 2014: 4), using popular media outlets to propagate antagonistic narratives that agitated latent political disputes and primed the historically Russian identity of these regions (O'Loughlin et al., 2017; Pakhomenko et al., 2018; Pomerantsev and Weiss, 2014). The use of compatriot organisations – funded by Russia to ostensibly maintain contact with Ukraine's large Russian diaspora – were also seen as essential methods for Russia to project antagonistic strategic narratives and stoke tensions endemic in the societies targeted (Grigas, 2016: 1). Such activities were seen as instrumental in steering discourse

in Donetsk, Luhansk, and Crimea, and mobilising segments of their populations towards the conflicts that later erupted (Kofman et al., 2017: 51).

Russia has also been accused of engaging in malign information influence in other central and eastern European regions. The Baltic States of Estonia, Latvia, and Lithuania are examples of regions thought to be targeted by Russian information operations (Król, 2017: 1). NATO's StratCom centre, among others, has warned of the increasing presence of aggressive Russian narratives in their information environments as Russia seeks closer integration with the states, and particularly, with their sizeable Russian-speaking minority communities (Bērziņa and Cepurītis, 2018: 26; Flanagan et al., 2019: 5). Russia's growing influence in the Western Balkans region has also been credited to information operations, where cultural sympathies and the low reach of government have meant that pro-Russian narratives have found fertile ground in local media outlets (Bechev, 2019; Galeotti, 2018; Salvo and De Leon, 2018: 1). Indeed, the region has been identified as a key battleground for Russia, which has a vested political and economic interest in keeping the countries comprised in the region in its sphere of influence (Galeotti, 2018: 2). Countries such as Bosnia and Herzegovina or Serbia have been highlighted as particularly susceptible to Russian malign information influence, with analysts suggesting the countries' complicated social, political, and economic environments create civilian domains that are particularly reactive to Russian antagonistic narratives (Zamfir, 2020: 36).

Moving west, national elections in the United States and across Europe have become noted targets of Russian information operations promoting antagonistic strategic narratives. These operations were intended to obstruct fair democratic processes by clouding various information environments and impeding citizens' ability to make informed decisions. Analyses of malign information influence surrounding the recent Italian and U.S. general elections (Flore et al., 2019; Kluver et al., 2019) or the UK Brexit referendum (Flaherty and Roselle, 2018) provide just a few examples that demonstrate these activities. Unexpected 'hot-button' societal events or issues, such as the ongoing Coronavirus pandemic (Colley et al., 2020) or the Black Lives Matter protests in the United States (Arif et al., 2018) have also been exploited by Russian malign information influence – warped opportunistically into weapons that sought to exacerbate tensions in different societies and further Russia's political agenda.

However, the threat of malign information influence is not limited to Russian conduct, and many hostile actors have been accused of engaging in such hostile activity. In 2019, Twitter published the first release from their archive of state-sponsored information operations on their platform. Here, they exposed states such as Saudi Arabia, Iran, and China and their use of fake news and 'bot armies' to distort and amplify contentious geopolitical narratives and the 'attack' the civilian domain of various target societies (Twitter, 2019). More recently, extensive Chinese information operations were exposed that, similarly to Russia, also sought to leverage societal events such as the Coronavirus pandemic or Black Lives Matter protests to sow discord in states, namely in Italy and the United States (Bechis, 2020; Factcheck.org, 2020).

Psychological research on malign information influence: What's there and what's missing?

Given that malign information influence concerns the manipulation of citizens' perceptions, a psychological perspective is very relevant. Indeed, there are decades of psychological research that provide indirectly relevant insights for how malign information influence might function. This ranges from, for example, understanding how narrative frames affect participation in social

movements to how news framing can affect citizen cognitions (Price et al., 1997). In terms of research with direct relevance to malign information influence, recent attention from psychology has mainly focused on investigating how we may address the threat that disinformation – a frequent component of antagonistic strategic narratives – poses. In this, psychology has examined the various cognitive mechanisms that enable belief in disinformation (Lewandowsky et al., 2013, 2017). Emotion research, for example, has shown that information that incites anger or anxiety can reduce one's reasoning capabilities, suggesting that more provocative false political information might increase belief and consequential political polarisation (Weeks, 2015). The degree of emotive reactiveness to false political information has also been linked to the likelihood of sharing this information with others, thus increasing its spread and potential impact (Vosoughi et al., 2018).

Cognitive processes that audiences use to decipher if information is false have also been explored. Different cognitive mechanisms have been used to explain why one may be deceived by false messaging if it is consistent with their pre-existing beliefs, if it is socially accepted, or if there is a lack of information regarding the message's source (Lewandowsky et al., 2012; Nisbet and Kamenchuk, 2019). Cognitive heuristics such as fluency or familiarity have been attributed to the increase in accessibility of false information in one's memory after repeated low-level exposure, such as on social media (Pennycook et al., 2018). More broadly, the experience of living in a post-truth world – a society where the objective standards of truth are no longer upheld – has been linked to psychological states such as learned helplessness, a state of cognitive exhaustion which is thought to further increase susceptibility to believing disinformation (Amichai-Hamburger et al., 2003; Dahlgren, 2018: 23; Nisbet and Kamenchuk, 2019: 72).

These psychological insights have had fundamental implications for policy development seeking to mitigate disinformation's impact. For example, psychologists have emphasised the utility of explicit pre-exposure warnings (Ecker et al., 2010), the ineffectiveness of rebuttals (Ecker, 2017), and the promise of education programmes teaching media literacy and reasoning skills (Lewandowsky et al., 2012) in reducing the impact of disinformation. These recommendations have led to the development of effective media literacy programmes for young adults (McDougall et al., 2018) or Facebook's, and later other social media platforms, addition of warnings to content flagged as containing potentially false information (Mosseri, 2016). A very recent example of how these insights have been integrated is Twitter's addition of labels to state-affiliated media accounts (Twitter, 2020). The notion of transparency regarding message sources is one that has been promoted by counter-propaganda experts (McGeehan, 2018: 55) and is in line with dual-process theories of persuasion such as the Elaboration Likelihood Model (Petty and Cacioppo, 1986) or the Heuristic-Systematic Model (Chaiken, 1980). Indeed, psychology has made a clear contribution to understanding *how* malign information influence may operate and measures governments or companies may take to alleviate the effects of disinformation. However, a crucial aspect of psychology's potential contribution to understanding malign information influence would be that of *effects*. What psychological effects does malign information influence elicit in citizens targeted by such activities? More specifically, how do antagonistic strategic narratives – which often include, but are not limited to, disinformation – shape citizen cognitions?

The erosion of trust in democracy, authorities, or fellow citizens – dubbed the 'trinity of trusts' (Ingram, 2020: 16) – and the incitement of different (positive or negative) emotions are commonly discussed psychological effects and frequently suggested as contributors to the destabilisation that malign information influence can precipitate in societies (Ducheine et al., 2017: 6; Flore, 2020: 13; Helmus et al., 2018: 10; Ingram, 2020: 12). Indeed, it is acknowledged that strategic narratives are

crafted 'in order to intentionally reach the aim of affecting an audience on a security issue, influencing and managing "expectations as they shape language and ideas"' (Livingston and Nassetta, 2018: 102 in Coticchia and Catanzaro, 2020: 10). NATO has outlined the cognitive and psychological domain as the 'most important, as it consists of cognition and emotions, which affect an individual's decision-making' (NATO, 2015: 1–2). Moreover, the aforementioned psychological research was conceived to reduce the ostensible pernicious effects of disinformation on citizens' perceptions, attitudes, and behaviours (Nisbet and Kamenchuk, 2019: 66). Yet, despite this apparent emphasis on prevention, only a small (albeit promising) amount of research has sought to establish evidence of psychological effects. However, the nature of these studies, being qualitative (for example, Crilley and Chatterje-Doody, 2020), correlational (for example, Bērziņa, 2018a, 2018b; Ingram, 2020), measured through proxies (for example, Barfar, 2019; Lanoszka, 2019; Nimmo et al., 2017) or inconclusive (for example, Bail et al., 2020), prevents deduction of any causal relationships between information influence activities and psychological effects.

This research gap has been highlighted in recent discourse. Wagnsson emphasises the necessity for an understanding of the effects of malign information influence and 'better knowledge about the seriousness of the perceived danger and also about what kinds of damage it can do' (2020: 14). Tucker and colleagues identify the lack of clarity regarding the psychological effects of exposure to information operations as a substantive research gap in our knowledge of online polarisation (2018: 57). Wallenius and Nilsson (2019) note the dearth of effect-based studies into such operations, showing that available research demonstrating effects in target populations is often dated and limited to case studies. Colley et al. highlight that policy decisions are typically based on assumptions about what effects information operations are having on attitudes, behaviours, and societal cohesion, rather than robust evidence of these effects – something impossible to measure without knowledge of the precise effects to be measured (2020: 92). Ultimately, statements regarding if and to what extent we can claim that malign information influence triggers deleterious psychological effects in target populations and if so, what these specific effects are, remain unsubstantiated by psychological research.

In line with other authors, we suggest that experimental psychological research would offer a promising avenue to better substantiate these statements (Colley et al., 2020: 123; Helmus et al., 2020: 1; Tucker et al., 2018: 71; Wagnsson, 2020: 14). The benefits of such an approach would be the accurate and quantitative measuring of the effects of exposure to antagonistic strategic narratives on different psychological factors. Randomised experimental research also allows for better control of any external or biasing factors that might introduce noise to the measurement of these relationships (Bryman, 2016). In general, this deductive method would allow more precision when ascertaining possible causal relationships. Such controlled environments mean we can assume that any psychological effects observed would likely be the effects of the exposure to a particular antagonistic strategic narrative. Obtaining these results would therefore offer more conclusive evidence of if malign information influence tactics elicit psychological effects and specificity regarding what these effects are. This evidence would consequently afford more certainty to the discourse surrounding the possible impact of malign information influence in the civilian domain.

The possibilities for experimental psychological research into malign information influence

In focusing on the psychological effects of malign information influence and the experimental possibilities to unpack them, we advance a relatively uncharted approach to examining malign

information influence by foreign hostile actors. The proposed research agenda is therefore fairly open, with several possible psychological effects to explore and a choice of methodologies to do so. In this next section, we will sketch some parameters within which this experimental research can be conducted, providing clarity and structure to the process of creating hypotheses and designing experimental paradigms. In doing so, we will explore how suggestions for both hypotheses and methodologies can be informed and inspired by research areas with high thematic proximity, such as research into the psychological effects of general conspiracy theories or populist rhetoric. These research areas are pertinent when considering malign information influence as the antagonistic strategic narratives projected by foreign hostile actors often incorporate analogous narrative devices. Russian state-funded media outlets *RT* (formerly Russia Today) and *Sputnik*, for example, are widely noted to weave elements of various conspiracy and populist narratives into their content (Chatterje-Doody and Crilley, 2019; Yablokov, 2015). Such research areas, therefore, provide a solid platform to base research that looks more at the specific context of malign information influence.

Selecting psychological factors

In beginning this experimental process, the first question one must ask themselves is which specific psychological factors do we predict antagonistic strategic narratives to affect, and in what ways? We suggest that the most relevant effects to be researched lie in two main psychological components: trust and (negative) emotions. As mentioned, both components are frequently referenced in discourse when discussing the deleterious effects of malign information influence activities. The erosion of trust – be that in the government and authorities, between citizens in a society, or in democracy itself – has been considered a vital goal for hostile actors aiming to harm target states through informational means (Ingram, 2020: 17). Similarly, the inciting of negative emotional states such as anger, fear, disgust, or confusion are often considered as destabilising psychological effects that result from such influence efforts in citizens (Flore, 2020; Wither, 2016: 83). Indeed, the aforementioned qualitative or correlational studies into psychological effects of Russian strategic narration have often centred on trust or emotional factors, such as Crilley and Chatterje-Doody's qualitative exploration of audiences' emotional responses to Russian media outlet RT's narratives of Russia's military intervention in Syria (2020) or Bērziņa's correlational examination of political trust and Russian media consumption in Latvia (2018a).

Their relevance is further underscored by their common observation in experimental research investigating the effects of conspiracy or populist narratives. Experimental conspiracy theory research has evidenced the detrimental effects that such narratives can have on trust in politicians (Douglas, 2017), in democracy (Jolley and Douglas, 2014), or between different societal groups (Jolley et al., 2020). Studies have also shown how depending on their content, various conspiracy narratives can stimulate higher levels of fear, anger, or disgust reactions in those exposed to them than those not (Jolley and Paterson, 2020: 629; Jolley et al., 2020). Similarly, experimental examinations of populist narratives routinely demonstrate their adverse effects on confidence in political and 'mainstream media' institutions or attitudes towards different 'threatening' societal 'outgroups' (Albertazzi and McDonnell, 2008; Hameleers, 2020; Hameleers and Fawzi, 2020; Schmuck and Matthes, 2015). Moreover, fear, anger, and to a lesser extent, disgust, have been centred as the emotional nexus of populist rhetoric (Rico et al., 2017; Salmela and von Scheve, 2017; Wirz, 2018), being the most common emotional responses to the (real or imagined) societal threats that are often depicted in populist narratives (Vasilopoulos et al., 2019: 681; Wirtz et al., 2016).

Generating theoretical models

To generate hypotheses about the possible effects of antagonistic strategic narratives on trust factors or negative emotions, it is crucial to consider them in the context of the broader strategic goals that hostile political actors have in the international arena. Knowledge of the macro-level intentions different actors have towards a targeted state – e.g. to weaken its societal cohesion or diminish the national government's legitimacy – can inform predictions of the micro-level effects that their information operations targeting this state might have. Connecting to extant analyses that consider the wider geopolitical functions of different narrative approaches (e.g. Flaherty and Roselle, 2018; Khaldarova and Pantti, 2016; Ramsay and Robertshaw, 2019) would be a cogent and productive step here. To illustrate this, let us briefly consider Wagnsson and Barzanje's (2019) analytical framework examining the geopolitical functions of *Sputnik*'s antagonistic narration of Sweden, and expanded later to the Nordic countries (Deverell et al., 2020). The proposed framework delineates three key strategies that each fulfil a particular strategic goal for the Kremlin, including 'destruction', described as a narrative strategy that aims to undermine perceptions of a target state's 'military resolve, economic strength, and governmentality' (Deverell et al., 2020: 4). In this, 'destruction' aggregates narratives that emphasise a state's internal divisions, military vulnerabilities, and 'incompetent' political leadership, and portray the state as an unreliable actor in the global political arena (Wagnsson and Barzanje, 2019: 13). If we were to predict the psychological effects of 'destruction' narratives in the domestic audience, then, one might hypothesise decreases in governmental trust or increases in fear as readers appraise the implications of this 'state weakness' information for their personal circumstances. Such an example, while short, demonstrates how drawing on granular analyses of hostile actors' strategic narratives and discussions of their strategic intent can enable contextually and theoretically-grounded hypotheses.

An important consideration when extrapolating these hypotheses would be the order of the psychological effects. Theoretically, it is reasonable to suggest that any destabilisation provoked in citizens by antagonistic strategic narratives is due to an interplay between different constellations of psychological effects. In this, a narrative that triggers an effect on a certain factor can lead to the triggering of consequent effects on other, related psychological factors. Consider a hypothetical example (illustrated in Figure 1) whereby exposure to an antagonistic narrative that emphasises internal divisions and the likelihood of domestic conflict within a state leads to increased levels of fear in exposed citizens. It is reasonable to submit that this increased fear was facilitated by an interim step, whereby citizens appraising what the information in this narrative means for their own personal safety are led to perceive a higher level of threat to their personal safety. The proposition would then be that this narrative led to an increase in fear, *through* the increase in perceived threat. This mechanism, whereby a third variable influences the relationship between two other variables, is statistically termed *mediation*. From a mediation perspective, we might posit that a single psychological factor, such as perceived threat, can behave as a central 'trunk' effect, with several 'branches' of second or third-order psychological effects, such as increased political or social distrust. We can test these constellations through forming falsifiable hypotheses, and these insights have profound benefits for countermeasure opportunities – an idea we will return to in a later section.

Previous research provides evidence that would support these mediation configurations of psychological effects. Increasing threat perceptions have often positioned as a mediator between exposure to negative media narratives and increased in prejudiced attitudes towards 'outgroups' (Duckitt, 2006; Matthes and Schmuck, 2017; Schmuck and Matthes, 2017). Increased disgust has

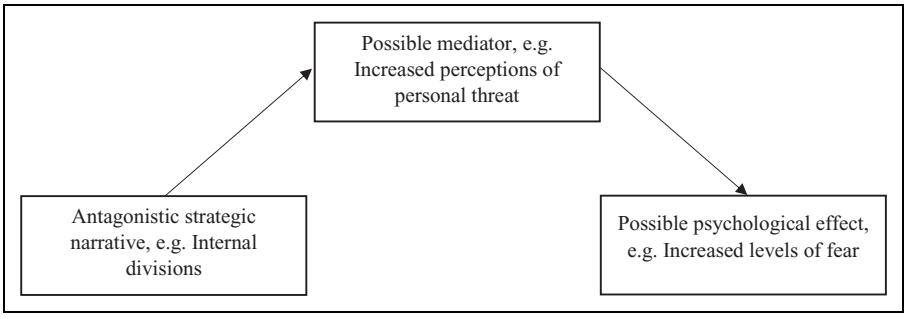

Figure 1. Diagram showing a hypothetical mediation between antagonistic strategic narratives, their psychological effects and potential psychological mediators.

also been shown as a mediating factor in increasing negative feelings towards societal 'outgroups' (Wirtz et al., 2016). Anger, fear, and enthusiasm have all been identified as important mediating variables in understanding the psychological responses to anti-immigration media narratives (Lecheler et al., 2015) and wider populist discourse (Rico et al., 2017). Meanwhile, dependency structures have also been observed between trust factors, whereby citizen's trust in the government has been evidenced as directly contingent on their trust in other citizens in their society (Tao et al., 2014). Moreover, comparable mechanisms have already been observed in studies investigating the effects of conspiracy theories. Exposure to intergroup conspiracy theories was shown to lead to a so-called 'secondary-transfer effect' whereby increased prejudice was felt towards unrelated outgroups, but only by increasing prejudice towards the outgroup that the conspiracy theory was based on (Jolley et al., 2020: 29). Such evidence suggests that careful attention should be placed on conceptualising the relationships between narrative and the cascading sequences of psychological effects that antagonistic narration can provoke.

A further consideration when testing these hypotheses would be to ask for whom the effects of exposure to certain antagonistic strategic narratives would be strongest. Societies are comprised of a broad patchwork of groups and citizens, each with its own unique psychological compositions. Knowledge of individual differences that amplify adverse psychological effects can therefore offer opportunities to assess and address specific vulnerabilities in society. Consider a variant of the earlier hypothetical example (also illustrated in Figure 2): exposure to a narrative emphasising internal societal divisions and suggesting to citizens that their country is approaching an unavoidable conflict seems likely to trigger greater feelings of fear. Yet, it seems also likely that this effect should be compounded in citizens already experiencing inherently high levels of anxiety about the future. This is a statistical mechanism called *moderation*, whereby the strength of a relationship between two variables depends on a third. Extant research has evidenced the importance of moderation effects on responses to political messaging, whereby different psychological characteristics robustly lead to different responses to certain political messages (Gerber et al., 2013). Moreover, counterterrorism research has emphasised the need for research into individual differences and their effects on the efficacy of different counter-narratives (Braddock and Horgan, 2016: 399). The lack of knowledge regarding potentially amplifying factors has been identified as a crucial knowledge gap in our understanding of the effects of information operations (Wallenius and Nilsson, 2019: 408). Experimental psychological research, once more, offers an encouraging pathway to address this knowledge gap.

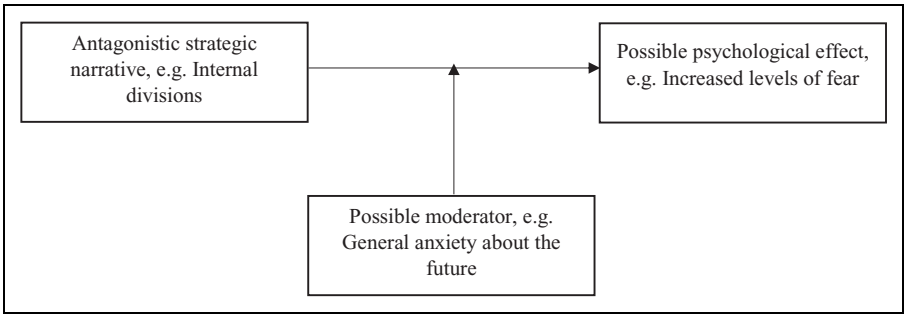

Figure 2. Diagram showing a hypothetical moderation between antagonistic strategic narratives, their psychological effects and potential psychological moderators.

Again, it is practical to draw on thematically-related psychological research in devising relevant suggestions for possible moderating psychological characteristics (summarised in Table 1). For example, recent research has evidenced the role of an individual's levels of self-efficacy, self-control, and legal cynicism in moderating violent intentions after exposure to conspiracy theories (Rottweiler and Gill, 2020). Similarly, studies have examined how feelings of powerlessness, uncertainty, and disillusionment can contribute to reducing political engagement or intentions to engage in pro-social behaviour after exposure to conspiracy theories (Jolley and Douglas, 2014). While this research suggests a mediation relationship – that these negative effects arise because the conspiratorial narratives induce these feelings of powerlessness, uncertainty, or disillusionment – it would also be reasonable to consider these psychological factors as more long-standing, stable traits that exist prior and independently to exposure to conspiracy narratives. Conversely, an individual's level of political cynicism (Bos et al., 2013), one's level of attachment to a national or international identity (Hameleers et al., 2017), and one's formal education level (Matthes and Schmuck, 2017; Schmuck and Matthes, 2015) have all been highlighted as moderators by research into the psychological effects of populist discourse.

Designing experimental methods

Having generated a theoretical model of the effects of antagonistic strategic narratives, the next step for researchers would be to devise the methodology that one should use to experimentally test these hypotheses. Of course, there are several research avenues that one might consider here. Longitudinal studies that track and monitor the psychological effects of exposure to these strategic narratives would offer key realistic insights into how malign information influence affects citizens. However, longitudinal designs can create difficulties when it comes to maintaining the strict experimental conditions necessary to draw causal conclusions. Informative experimental paradigms also stem from those that allow for the direct comparison of the levels of the psychological factors in question, varied between participants who were exposed or not exposed to antagonistic narrative content. Turning again to extant literature investigating the psychological effects of different narratives, we can see that such between-subject paradigms are by far the most popular experimental paradigm examining such effects. This design has been regularly employed to demonstrate causal relationships between different conspiratorial narratives (e.g. Jolley and

Table 1. Potential moderators identified by thematically-relevant experimental studies.

Potential moderator(s)	Related literature
Self-efficacy, self-control, and legal cynicism	Rottweiler and Gill (2020)
Powerlessness, uncertainty, and disillusionment	Jolley and Douglas (2014)
Political cynicism	Bos et al. (2013)
Attachment to national or international identity	Hameleers et al. (2017)
Formal education	Schmuck and Matthes (2015) and Matthes and Schmuck (2017)

Douglas, 2014; Jolley et al., 2020) or types of pernicious media discourse (e.g. Lecheler et al., 2015; Shen et al., 2014) and their consequent psychological effects.

To distil the materials one could use in these experimental studies, one can again draw inspiration from granular analyses of different information operations. Qualitative content analyses of the narratives projected by the hostile actor in question are one method that a researcher might use to devise these materials. Qualitative analyses can extract the core narrative mechanisms of a hostile actor, which can then be readily assimilated into any experimental stimuli. A recent study by Hameleers (2020) provides an example of how insights from qualitative analyses of narrative content can be directly transposed into the design of manipulations for experimental research. Suggestions might also be taken from several recent examinations of the most frequent frames, strategic narratives, or master narratives used by different hostile actors' information operations, including China (Ford, 2015; Kluver et al., 2019), Iran (Cooley et al., 2019) or Russia (Levinger, 2018; Nimmo et al., 2017; Ramsay and Robertshaw, 2019; Rebegea, 2019). Overviews of the general narrative trends of different hostile actors provide useful reference points to ensure the contextual appropriateness of any experimental stimuli designed.

Limitations and difficulties of experimental research

It is wise to acknowledge that there are, of course, considerable methodological hurdles to overcome when devising experimental research into the effects of malign information operations on citizens (Tucker et al., 2018: 71; Wallenius and Nilsson, 2019: 407). Ethical concerns that arise when seeking to study any pernicious psychological effects in participants would have to be addressed. There is also an inherent trade-off between specificity and generalisability in this research: well-devised experiments that cleanly delineate psychological effects of a certain hostile actor's specific and contextualised narrative strategies will suffer from a lack of generalisability to other hostile actors. Conversely, more generalised investigations into the mechanisms will be limited in the precision that their insights would offer. This balance between generalisability and precision is not specific to this research direction but is something any researcher embarking on this experimental research would have to weigh for themselves, depending on the specific questions that they seek to investigate.

More broadly, it is wise to be realistic about the conclusions one might be able to draw from the proposed experimental research. Experimental research is a fruitful endeavour for those seeking to provide clarity and insight into if causal relationships between stimuli and their psychological effects exist. Yet, how far any effects evidenced can be extrapolated to real life, where exposure to strategic narratives is natural, repetitive, and self-directed, is uncertain. Part of this stems from the

difficulty in capturing the specific audience who are exposed to such narrative content. Limited research into the target audience of, for example, Russian foreign language state-media outlets suggests that the audience can often be characterised by specific subsets of pre-existing attitudes (Müller and Schulz, 2019). Such intricacies are unlikely to be represented fully in the samples feasibly acquired for experimental research. While aforementioned moderation analyses can broach this concern it is likely to remain a significant limitation when drawing conclusions. A similar concern would be the inability of experimental research to recreate the low-level exposure that occurs for citizens exposed to strategic narrative content in real life. Some studies have attempted just this – Schmuck and Tribastone (2020) examined the effects of exposure to anti-Islamic political discourse longitudinally. However, it is difficult to maintain experimental conditions in such repeated-exposure designs (Rooduijn et al., 2017: 142). These are long-standing difficulties that psychological and communication science have contended with for decades, and they form critical concerns that any researchers broaching this research agenda should be aware of and incorporate into any conclusions.

The possibilities of experimental psychological research into malign information influence

Having discussed a clear agenda for how we might acquire evidence of the possible psychological effects of malign information influence, we will now explain what these insights can contribute to different countermeasure efforts. In recent years, addressing the complex security environment, and particularly the rising threat of information operations to civil society, have become prolific topics in international security discourse. Countermeasures that have been implemented have included strategic communication efforts seeking to control and reshape the images and narratives surrounding political institutions, campaigns to increase awareness of disinformation, and policy responses aimed at improving online regulation and fact-checking (European Commission, 2016: 2; Giumelli et al., 2017: 147). The proliferation of such efforts has, however, been met with some criticisms. For example, modern strategic communication efforts have been critiqued for failing to integrate germane psychological insights (Payne, 2011: 6) and as mentioned before, policy responses have been criticised for resting on assumptions of the effects of malign influence activities instead of on firm, empirical evidence (Colley et al., 2020: 92). Experimental research that validates the precise psychological effects of information operations would contribute to addressing these criticisms, and we believe, would confer valuable insights to potential countermeasures or policy developments.

Options for impact assessments

The first opportunity that experimental research into psychological effects offers lies in the precision it would afford researchers seeking to understand the impact of malign information influence. Currently, longitudinal surveys monitoring broad-scale public opinion are used to track the potential impact of malign information influence (see, for example, Bērziņa, 2018a). Validated knowledge of the specific first or second-order psychological effects that various antagonistic strategic narratives can have means subsequent studies assessing impact can do so with greater accuracy. Evidence that, for example, a key Russian approach is to leverage various political controversies to undermine the competency of the European Union, and that these antagonistic narratives can have significant effects on psychological factors such as increasing feelings of

frustration in citizens, means that surveys or opinion-gathering tools can be refined accordingly. This knowledge can also contribute important insights for predicting and anticipating the impact of information operations in hypothetical scenarios. Specifically, evidence of possible psychological effects of antagonistic strategic narratives can complement computational methods (e.g. dynamical systems or agent-based models) that can be used to simulate how different antagonistic strategic narratives might emerge after hypothetical societal events (see, for example, Schmid et al., 2017 for a proposed framework). Empirically-backed insights into the psychological effects can then augment insights drawn from these simulations, whereby it would not only be possible to forecast the type and spread of these antagonistic narratives that might emerge, but also the potential societal impact of these antagonistic narratives in anticipated political situations.

Reactivity analyses

A second benefit of this research would lie in the possibility to generate analyses measuring reactivity to different antagonistic strategic narratives across societies or groups within a society. As explained above, the motivation for any experimental moderation analyses would be to identify any individual differences that amplify undesirable psychological effects. This presents the possibility to measure these individual differences in citizens through broad cross-cultural surveys or self-report studies. With these measurements, the average levels of these moderating traits or factors of interest can be assessed across societies or groups of interest and then amalgamated into indices of societal or group-level reactivity. These indices would be crude, but they can give broad directions about which societies, or which groups in societies, might be worth paying closer attention to given certain antagonistic narratives or certain political events. This type of societal index research echoes previous initiatives by counter-information operation institutions. The NATO Centre for Strategic Communication has worked on indices of society permeability, economic permeability, and political permeability towards Russian malign information influence in the Western Balkans region (Zamfir, 2020). Similarly, the GLOBSEC Policy Institute has published their index of societal vulnerability to subversive Russian influence in the Czech Republic, Slovakia, Poland, and Hungary (Milo and Klingová, 2017).

Targeted public initiatives, education, and campaigns

A deeper understanding of the complex, dependent structures of any psychological effects and individual differences that amplify these psychological effects means specific aspects of the civilian domain can be flagged that states can endeavour to reinforce before they are pursued by hostile actors. Evidence that, for example, governmental trust is a 'central' effect of certain antagonistic narratives, with second-order effects of reductions in trust between societal groups and increases in frustration, would give empirical validation that an intervention seeking to address this lack of political trust would be the most productive first step in developing deterrence efforts. This idea of improving societal resilience is a key element of emerging population-centric approaches to countering hybrid or grey-zone activities, such as 'modern deterrence' (Braw, 2019: 4), inspired by the Swedish 'Total Defence' model enacted during the Cold War (Wither, 2020: 62). As resilience initiatives, such as education schemes and outreach programmes, are often accompanied by large financial burdens, an awareness of the potential 'targets' in a society's wall would allow for more efficiency in deterrence efforts. By focusing deterrence efforts on these specific and validated vulnerabilities, it would be more difficult for hostile actors to exploit the

civilian domain and an adversary should view such endeavours as disproportionally and unreasonably effortful, creating successful resilience. This is a notion echoed by the European Centre of Excellence for Countering Hybrid Threats, who suggest that potential adversaries 'would refrain from taking aggressive actions [...] if the perception is that the costs of the attack would supersede the benefits' (Sørensen and Bach Nyemann, 2018: 2).

Smarter construction of counter-narratives

A final benefit lies in the opportunity for the smart construction of counter-narratives. Knowledge regarding the psychological effects of certain hostile actor's common narrative strategies and their effects means counter-narratives can be more efficiently built. Understanding how a hostile actor is likely to distort a societal event to enable their overarching narrative strategies affords the benefit of pre-emption to potential victims of these attacks. To use the previously mentioned hypothetical example, communication that seeks to address an overarching anti-EU narrative and its proposed negative effect on political trust in a certain societal group, can be prepared before any potential Russian reaction (Rebegea, 2019: 8). If therefore, a societal event occurs that seems likely to be distorted into enabling this narrative strategy, counter-narratives might be prepared that: offer more transparency about the EU's role in said event, reveal inconsistencies in the antagonistic narratives, and avoids accidental reinforcement of the antagonistic narrative's themes. This idea of using evidence of the psychological effects of hostile narratives to enhance counter-narrative construction is one already embraced by the field of counterterrorism to reduce support for violent extremism (e.g. Braddock and Horgan, 2016). The pre-emptory advantage means strategic communication can be troubleshot, refined, and then disseminated quickly and efficiently to elicit maximally defensive effects.

Conclusion

This article has sought to outline the mounting challenge malign information influence activities pose to international security bodies, as well as how psychological science has, thus far, contributed knowledge in how we may counter this challenge. We have then outlined how despite the commonly accepted notion that adversaries use malign information influence to elicit psychological effects in citizens within societies and destabilise the civilian domain, little empirical research has been done that demonstrates this. We have gone on to show how experimental psychological methodologies hold great promise in offering key insights regarding these effects. This includes the potential to offer more conclusive evidence of psychological effects caused by different antagonistic narratives and the identification of any psychological characteristics that might amplify these effects. Finally, we have advanced four main contributions that this research might make in terms of countering malign information influence. These were more precise measurement opportunities to assess the impact of information operations, the possibility to develop indices of reactivity to antagonistic narrative strategies, more specificity in creating deterrence initiatives, and insights that can be used to construct smarter strategic communication. We hope that this article has demonstrated how novel, empirically-based insights into the psychological effects of malign information influence can be gathered and the promising implications these insights have for those seeking to counter it.

Acknowledgement

We thank the reviewers for their insightful comments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by TNO and the MoD/FMW.

References

Albertazzi D and McDonnell D (2008) Twenty-first century populism: the spectre of western European democracy. In: Albertazzi D and McDonnell D (eds) *Twenty-First Century Populism: The Spectre of Western European Democracy*. Basingstoke: Palgrave Macmillan, pp. 1–11.

Almäng J (2019) War, vagueness and hybrid war. *Defence Studies* 19(2): 189–204.

Ambrosio T (2016) The rhetoric of irredentism: the Russian Federation's perception management campaign and the annexation of Crimea. *Small Wars and Insurgencies* 27(3): 467–490.

Amichai-Hamburger Y, Mikulincer M and Zalts N (2003) The effects of learned helplessness on the processing of a persuasive message. *Current Psychology* 22: 37–46.

Antoniades A, Miskimmon A and O'Loughlin B (2010) *Great Power Politics and Strategic Narratives*. Brighton: The Centre for Global Political Economy.

Arif A, Stewart LG and Starbird K (2018) Acting the part: examining information operations within #BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 2: 1–27.

Bachmann S-D and Gunneriusson H (2015) Russia's hybrid warfare in the east: the integral nature of the information sphere. *Georgetown Journal of International Affairs* 16: 198–211. Available at: https://ssrn.com/abstract=2670527 (accessed 5 November 2020).

Bail CA, Guay B, Maloney E, et al. (2020) Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proceedings of the National Academy of Sciences of the United States of America* 117(1): 243–250. DOI: 10.1073/pnas.1906420116

Barfar A (2019) Cognitive and affective responses to political disinformation in Facebook. *Computers in Human Behavior* 101: 173–179.

Bechev D (2019) *Russia's Strategic Interests and Tools of Influence in the Western Balkans*. Riga: NATO Strategic Communications Centre of Excellence. Available at: https://stratcomcoe.org/russias-strategic-interests-and-tools-influence-western-balkans (accessed 5 November 2020).

Bechis F (2020) Così Russia e Cina fanno propaganda in Italia. Il rapporto del Copasir. *Formiche*. Available at: https://formiche.net/2020/05/russia-cina-propaganda-italia-rapporto-copasir/ (accessed 5 November 2020).

Bērziņa I (2018a) Political trust and Russian media in Latvia. *Journal on Baltic Security* 4(2): 1–8.

Bērziņa I (2018b) Russia's narratives and public opinion in the Baltic States, Finland, and Sweden. In: Lange-Ionatamišvili E (ed.) *Russia's Footprint in the Nordic-Baltic Information Environment*. Riga: NATO Strategic Communications Centre of Excellence, pp. 79–103. Available at: https://www.stratcomcoe.org/russias-footprint-nordic-baltic-information-environment-0 (accessed 5 November 2020).

Bērziņa I and Cepurītis M (2018) Russia's grand strategy and its implications on the information environment of the Nordic-Baltic region. In: Lange-Ionatamišvili E (ed.) *Russia's Footprint in the Nordic-Baltic*

Information Environment. Riga, pp. 16–31. Available at: https://www.stratcomcoe.org/russias-footprint-nordic-baltic-information-environment-0 (accessed 5 November 2020).

Betz D (2008) The virtual dimension of contemporary insurgency and counterinsurgency. *Small Wars and Insurgencies* 19(4): 510–540.

Bos L, Van Der Brug W and De Vreese CH (2013) An experimental test of the impact of style and rhetoric on the perception of right-wing populist and mainstream party leaders. *Acta Politica* 48(2): 192–208.

Braddock K and Horgan J (2016) Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict and Terrorism* 39(5): 381–404.

Braw E (2019) *Re-Thinking Deterrence*. Centre for Historical Analysis and Conflict Research. Available at: https://chacr.org.uk/wp-content/uploads/2020/01/20190201_Issue16_CHACR_GAP_Briefing_Re_Thinking_Deterrence.pdf (accessed 5 November 2020).

Bressan S and Sulg M-L (2020) Welcome to the grey zone: future war and peace. *New Perspectives* 28(3): 379–397.

Bryman A (2016) *Social Research Methods*. Oxford: Oxford University Press.

Chaiken S (1980) Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39(5): 752–766.

Chatterje-Doody PN and Crilley R (2019) Populism and contemporary global media: populist communication logics and the co-construction of transnational identities. In: Stengel FA, MacDonald DB and Nabers D (eds) *Populism and World Politics*. Cham: Palgrave Macmillan, pp. 73–99.

Colley T (2017) Is Britain a force for good? Investigating British citizens' narrative understanding of war. *Defence Studies* 17(1): 1–22.

Colley T, Granelli F and Althuis J (2020) Disinformation's societal impact: Britain, Covid, and beyond. *Defense Strategic Communications* 8: 89–140.

Cooley S, Hinck R, Kitsch S, et al. (2019) *Iranian Media Narrative Perspectives on the US and Global Competition*. Media Ecology and Strategic Analysis (MESA). Available at: https://nsiteam.com/social/wp-content/uploads/2019/08/MESA-Group-Report_Iranian-Media-Narratives-v2.pdf (accessed 5 November 2020).

Coticchia F and Catanzaro A (2020) The fog of words: assessing the problematic relationship between strategic narratives, (master) frames and ideology. *Media, War and Conflict*. DOI: 10.1177/1750635220965622

Crilley R and Chatterje-Doody PN (2020) Emotions and war on YouTube: affective investments in RT's visual narratives of the conflict in Syria. *Cambridge Review of International Affairs* 33(5): 713–733.

Dahlgren P (2018) Media, knowledge and trust: the deepening epistemic crisis of democracy. *Javnost* 25(1–2): 20–27.

Deverell E, Wagnsson C and Olsson EK (2020) Destruct, direct and suppress: Sputnik narratives on the Nordic countries. *Journal of International Communication*. Epub ahead of print 09 September. DOI: 10.1080/13216597.2020.1817122

Douglas KM (2017) You just can't trust 'em: conspiracy theories erode people's faith in politicians and democracy itself. *International Politics and Society*. Available at: https://ips-journal.eu (accessed 5 November 2020).

Ducheine P (2014) Non-kinetic capabilities: complementing the kinetic prevalence to targeting. In: Ducheine P, Schmitt M and Osinga F (eds) *Targeting: The Challenges of Modern Warfare*. The Hague: T.M.C. Asser Press, pp. 201–230.

Ducheine PAL, van Haaster J and van Harskamp R (2017) Manoeuvring and generating effects in the information environment. In: Ducheine PAL and Osinga FPB (eds) *Netherlands Annual Review of Military Studies 2017*. The Hague: T.M.C. Asser Press, pp. 155–179.

Duckitt J (2006) Differential effects of right wing authoritarianism and social dominance orientation on outgroup attitudes and their mediation by threat from and competitiveness to outgroups. *Personality and Social Psychology Bulletin* 32(5): 684–696.

Ecker UKH (2017) Why rebuttals may not work: the psychology of misinformation. *Media Asia* 44(2): 79–87.

Ecker UKH, Lewandowsky S and Tang DTW (2010) Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory and Cognition* 38: 1087–1100.

European Commission (2016) *Joint Framework on Countering Hybrid Threats: A European Union Response*. Brussels: European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016JC0018 (accessed 5 November 2020).

Factcheck.org (2020) Falsehoods follow George Floyd's death. *Factcheck.org*. Available at: https://www.factcheck.org/2020/05/falsehoods-follow-george-floyds-death/ (accessed 5 November 2020).

Flaherty E and Roselle L (2018) Contentious narratives and Europe: conspiracy theories and strategic narratives surrounding RT's Brexit news coverage. *Journal of International Affairs* 71(1): 53–61.

Flanagan S, Osburg J, Binnendijk A, et al. (2019) *Deterring Russian Aggression in the Baltic States Through Resilience and Resistance*. Santa Monica, CA: RAND Corporation.

Flore M (2020) *Understanding Citizens' Vulnerabilities (II): From Disinformation to Hostile Narratives*. Luxembourg: European Commission.

Flore M, Balahur A, Podavini A, et al. (2019) *Understanding Citizens' Vulnerabilities to Disinformation and Data-Driven Propaganda Case Study: The 2018 Italian General Election*. Luxembourg: European Commission.

Ford C (2015) Puncturing Beijing's propaganda bubble: seven themes. *New Paradigms Forum*. Available at: http://www.newparadigmsforum.com/NPFtestsite/?m=201511 (accessed 5 November 2020).

Freedman L (2006) Networks, culture and narratives. *The Adelphi Papers* 45(379): 11–26.

Galeotti M (2018) *Do the Western Balkans Face a Coming Russian Storm?* London: European Council on Foreign Relations. Available at: http://www.ecfr.eu/publications/summary/do_the_western_balkans_face_a_coming_russian_storm (accessed 5 November 2020).

Gerber AS, Huber GA, Doherty D, et al. (2013) Big Five personality traits and responses to persuasive appeals: results from voter turnout experiments. *Political Behavior* 4(35): 687–728.

Giumelli F, Cusumano E and Besana M (2017) From strategic communication to sanctions: the European Union's approach to hybrid threats. In: Cusumano E and Corbe M (eds) *A Civil-Military Response to Hybrid Threats*. Cham: Palgrave Macmillan, pp. 145–167.

Grigas A (2016) *How Soft Power Works: Russian Passportization and Compatriot Policies Paved Way for Crimean Annexation and War in Donbas*. Available at: https://www.atlanticcouncil.org/blogs/ukrainealert/how-soft-power-works-russian-passportization-and-compatriot-policies-paved-way-for-crimean-annexation-and-war-in-donbas/ (accessed 5 November 2020).

Hameleers M (2020) They are selling themselves out to the enemy! The content and effects of populist conspiracy theories. *International Journal of Public Opinion Research*. Epub ahead of print 17 March. DOI: 10.1093/ijpor/edaa004

Hameleers M, Bos L and de Vreese CH (2017) "They did it": the effects of emotionalized blame attribution in populist communication. *Communication Research* 44(6): 870–900.

Hameleers M and Fawzi N (2020) Widening the divide between them and us? The effects of populist communication on cognitive and affective stereotyping in a comparative European setting. *Political Communication* 37(5): 612–634.

Helmus T, Bodine-Baron E, Radin A, et al. (2018) *Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe*. Santa Monica, CA: RAND Corporation.

Helmus TC, Marrone JV, Posard MN, et al. (2020) *Russian Propaganda Hits Its Mark: Experimentally Testing the Impact of Russian Propaganda and Counter-Interventions*. Santa Monica, CA: RAND Corporation.

Ingram HJ (2020) The strategic logic of state and non-state malign 'influence activities': polarising populations, exploiting the democratic recession. *RUSI Journal* 165(1): 12–24.

Jolley D and Douglas KM (2014) The social consequences of conspiracism: exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology* 105(1): 35–56.

Jolley D, Meleady R and Douglas KM (2020) Exposure to intergroup conspiracy theories promotes prejudice which spreads across groups. *British Journal of Psychology* 111(1): 17–35.

Jolley D and Paterson JL (2020) Pylons ablaze: examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology* 59(3): 628–640.

Khaldarova I and Pantti M (2016) Fake news: the narrative battle over the Ukrainian conflict. *Journalism Practice* 10(7): 891–901.

Kluver R, Cooley S and Hinck R (2019) Contesting strategic narratives in a global context: the world watches the 2016 U.S. election. *International Journal of Press/Politics* 24(1): 92–114.

Kofman M, Migacheva K, Nichiporuk B, et al. (2017) *Lessons from Russia's Operations in Crimea and Eastern Ukraine*. Santa Monica, CA: RAND Corporation.

Król A (2017) *Russian Information Warfare in the Baltic States – Resources and Aims*. Warsaw: The Warsaw Institute Review. Available at: https://warsawinstitute.org/russian-information-warfare-baltic-states-resources-aims/ (accessed 5 November 2020).

la Cour C (2020) Theorising digital disinformation in international relations. *International Politics* 57: 704–723.

Lanoszka A (2016) Russian hybrid warfare and extended deterrence in eastern Europe. *International Affairs* 92(1): 175–195.

Lanoszka A (2019) Disinformation in international politics. *European Journal of International Security* 4(2): 227–248.

Lecheler S, Bos L and Vliegenthart R (2015) The mediating role of emotions: news framing effects on opinions about immigration. *Journalism and Mass Communication Quarterly* 92(4): 812–838.

Levinger M (2018) Master narratives of disinformation campaigns. *Journal of International Affairs* 71(1.5): 43–52. Available at: https://www.jstor.org/stable/e26508110 (accessed 5 November 2020).

Lewandowsky S, Ecker UKH and Cook J (2017) Beyond misinformation: understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition* 6(4): 353–369.

Lewandowsky S, Ecker UKH, Seifert CM, et al. (2012) Misinformation and its correction. *Psychological Science in the Public Interest* 13(3): 106–131.

Lewandowsky S, Stritzke WGK, Freund AM, et al. (2013) Misinformation, disinformation, and violent conflict: from Iraq and the "war on terror" to future threats to peace. *American Psychologist* 68(7): 487–501.

McDougall J, Zezulkova M, van Driel B, et al. (2018) *Teaching Media Literacy in Europe: Evidence of Effective School Practices in Primary and Secondary Education*. Bournemouth: Network of Experts on Social Dimension of Education and Training (NESET).

McGeehan TP (2018) Countering Russian disinformation. *Parameters* 48(1): 49–58.

Matthes J and Schmuck D (2017) The effects of anti-immigrant right-wing populist ads on implicit and explicit attitudes: a moderated mediation model. *Communication Research* 44(4): 556–581.

Milo D and Klingová K (2017) *Vulnerability Index: Subversive Russian Influence in Central Europe*. Bratislava: GLOBSEC Policy Institute. Available at: https://www.globsec.org/publications/vulnerability-index-subversive-russian-influence-central-europe/ (accessed 5 November 2020).

Miskimmon A, O'Loughlin B and Roselle L (2014) *Strategic Narratives: Communication Power and the New World Order*. Abingdon: Routledge.

Mosseri A (2016) Addressing hoaxes and fake news. *Facebook Newsroom*. Available at: https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/ (accessed 5 November 2020).

Müller P and Schulz A (2019) Alternative media for a populist audience? Exploring political and media use predictors of exposure to Breitbart, Sputnik, and Co. *Information, Communication & Society* 24(2): 277–293.

NATO (2015) *Allied Joint Doctrine for Information Operations*. Brussels: RAND.

Nimmo B, Barojan D and Aleksejeva N (2017) *Russian Narratives on NATO's Deployment*. Digital Forensic Research Lab. Available at: https://medium.com/dfrlab/russian-narratives-on-natos-deployment-616e19c3d194 (accessed 5 November 2020).

Nisbet EC and Kamenchuk O (2019) The psychology of state-sponsored disinformation campaigns and implications for public diplomacy. *The Hague Journal of Diplomacy* 14(1–2): 65–82.

O'Loughlin J, Toal G and Kolosov V (2017) The rise and fall of "Novorossiya": examining support for a separatist geopolitical imaginary in southeast Ukraine. *Post-Soviet Affairs* 33(2): 124–144.

Pakhomenko S, Tryma K and J'moul F (2018) The Russian–Ukrainian war in Donbas: historical memory as an instrument of information warfare. In: Sayapin S and Tsybulenko E (eds) *The Use of Force Against Ukraine and International Law*. The Hague: T.M.C. Asser Press, pp. 297–312.

Payne K (2011) The psychology of strategic communication: some pointers. In: *KCL Insurgency Research Group and CIWAG US Naval War College Conference 'Strategic communications: the cutting edge'*, 10 May 2011.

Pennycook G, Cannon TD and Rand DG (2018) Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General* 147(12): 1865–1880.

Petty RE and Cacioppo JT (1986) The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology* 19: 123–205.

Pomerantsev P and Weiss M (2014) *The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money*. New York, NY: Institute of Modern Russia. Available at: https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev__The_Menace_of_Unreality.pdf (accessed 5 November 2020).

Price V, David T and Powers E (1997) Switching trains of thought: the impact of news frames on readers' cognitive responses. *Communication Research* 24(5): 481–506.

Ramsay G and Robertshaw S (2019) *Weaponising News. RT, Sputnik and Targeted Disinformation*. London: Kings College London. Available at: https://www.kcl.ac.uk/policy-institute/assets/weaponising-news.pdf (accessed 5 November 2020).

Rebegea C (2019) "Question more" – but not too much. Mapping Russia's malign master narratives in central and eastern Europe. In: Sultănescu D (ed.), *Challenges in Strategic Communication and Fighting Propaganda in Eastern Europe*. Amsterdam: IOS Press.

Rico G, Guinjoan M and Anduiza E (2017) The emotional underpinnings of populism: how anger and fear affect populist attitudes. *Swiss Political Science Review* 23(4): 444–461.

Rietjens S (2019) Unraveling disinformation: the case of Malaysia Airlines flight MH17. *International Journal of Intelligence, Security, and Public Affairs* 21(3): 195–218.

Rooduijn M, van der Brug W, de Lange SL, et al. (2017) Persuasive populism? Estimating the effect of populist messages on political cynicism. *Politics and Governance* 5(4): 136–145.

Roselle L, Miskimmon A and O'Loughlin B (2014) Strategic narrative: a new means to understand soft power. *Media, War and Conflict* 7(1): 70–84.

Rottweiler B and Gill P (2020) Conspiracy beliefs and violent extremist intentions: the contingent effects of self-efficacy, self-control and law-related morality. *Terrorism and Political Violence*. Epub ahead of print 20 October. DOI: 10.1080/09546553.2020.1803288

Salmela M and von Scheve C (2017) Emotional roots of right-wing political populism. *Social Science Information* 56(4): 567–595.

Salvo D and De Leon S (2018) Russia's efforts to destabilize Bosnia and Herzegovina. *German Marshall Fund of the United States*. Available at: https://www.jstor.org/stable/resrep18769 (accessed 5 November 2020).

Schmid KA, Züfle A, Pfoser D, et al. (2017) Predicting the evolution of narratives in social media. In: Gertz M, Renz M, Zhou X, Hoel E, Ku W-S, Voisard A, Zhang C, Chen H, Tang L, Huang Y, Lu C-T and Ravada S (eds) *Proceedings of the 2017 International Symposium on Spatial and Temporal Databases*. New York, NY: Springer, pp. 388–392.

Schmuck D and Matthes J (2015) How anti-immigrant right-wing populist advertisements affect young voters: symbolic threats, economic threats and the moderating role of education. *Journal of Ethnic and Migration Studies* 41(10): 1577–1599.

Schmuck D and Matthes J (2017) Effects of economic and symbolic threat appeals in right-wing populist advertising on anti-immigrant attitudes: the impact of textual and visual appeals. *Political Communication* 34(4): 607–626.

Schmuck D and Tribastone M (2020) Muslims take action. How exposure to anti-Islamic populist political messages affects young Muslims' support for collective action: a longitudinal experiment. *Political Communication* 37(5): 635–655.

Shen F, Ahern L and Baker M (2014) Stories that count: influence of news narratives on issue attitudes. *Journalism and Mass Communication Quarterly* 91(1): 98–117.

Sørensen H and Bach Nyemann D (2018) *Going beyond resilience: A revitalized approach to countering hybrid threats*. Hybrid CoE Strategic Analysis 13. Helsinki: Hybrid CoE. Available at: https://www.hybridcoe.fi/wp-content/uploads/2020/07/Strategic-analysis-13-Sorensen-Nyeman.pdf (accessed 11 February 2021).

Starbird K, Arif A and Wilson T (2019) Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–26.

Tao R, Yang DL, Li M, et al. (2014) How does political trust affect social trust? An analysis of survey data from rural China using an instrumental variables approach. *International Political Science Review* 35(2): 237–253.

Tucker JA, Guess A, Barberá P, et al. (2018) *Social Media, Political Polarization, and Political Disinformation*. Menlo Park, CA: Hewlett Foundation.

Twitter (2020) New labels for government and state-affiliated media accounts. Available at: https://blog.twitter.com/en_us/topics/product/2020/new-labels-for-government-and-state-affiliated-media-accounts.html (accessed 5 November 2020).

Twitter Safety (2019) Disclosing new data to our archive of information operations. Available at: https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019.html (accessed 5 November 2020).

Vasilopoulos P, Marcus GE, Valentino NA, et al. (2019) Fear, anger, and voting for the far right: evidence from the November 13, 2015 Paris terror attacks. *Political Psychology* 40(4): 679–704.

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.

Wagnsson C (2020) What is at stake in the information sphere? Anxieties about malign information influence among ordinary Swedes. *European Security* 29(4): 397–415.

Wagnsson C and Barzanje C (2019) A framework for analysing antagonistic narrative strategies: a Russian tale of Swedish decline. *Media, War and Conflict*. Epub ahead of print 8 November. DOI: 10.1177/1750635219884343

Wallenius C and Nilsson S (2019) A lack of effect studies and of effects: the use of strategic communication in the military domain. *International Journal of Strategic Communication* 13(5): 404–417.

Webb A (2014) *Britain, the BBC World Service and the Cold War*. London: Bloomsbury.

Weeks BE (2015) Emotions, partisanship, and misperceptions: how anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication* 65(4): 699–719.

Wilson T, Zhou K and Starbird K (2018) Assembling strategic narratives. *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): 1–26.

Wirtz C, Van Der Pligt J and Doosje B (2016) Negative attitudes toward Muslims in the Netherlands: the role of symbolic threat, stereotypes, and moral emotions. *Peace and Conflict* 22(1): 75–83.

Wirz DS (2018) Persuasion through emotion? An experimental test of the emotion-eliciting nature of populist communication. *International Journal of Communication* 12: 25.

Wither JK (2016) Making sense of hybrid warfare. *Connections: The Quarterly Journal* 15(2): 73–87.

Wither JK (2020) Back to the future? Nordic total defence concepts. *Defence Studies* 20(1): 61–81.

Yablokov I (2015) Conspiracy theories as a Russian public diplomacy tool: the case of *Russia Today* (*RT*). *Politics* 35(3–4): 301–315.

Zamfir R (2020) *Risks and Vulnerabilities in the Western Balkans*. Riga: NATO Strategic Communications Centre of Excellence. Available at: https://stratcomcoe.org/risks-and-vulnerabilities-western-balkans (accessed 5 November 2020).

Author biographies

Aiden Hoyle is a PhD candidate at the University of Amsterdam, TNO: Defence, Safety and Security and the Netherlands Defence Academy. His research focuses on the psychological dimension of hybrid warfare, and specifically, malign information activities.

Helma van den Berg is a Senior Human Behaviour Specialist at TNO Defence, Security, and Safety. Helma works on analyzing and changing behavior in the safety and security domain.

Bertjan Doosje is an Associate Professor of Psychology at the University of Amsterdam. He is interested in intergroup relations from a social and cultural psychological perspective.

Martijn Kitzen is an Associate Professor of War Studies at the Netherlands Defence Academy. Martijn focuses on irregular warfare and special operations and holds a particular interest in embedding population-centric approaches in military operations.