

PUBLICATIONS OF THE PRIME MINISTER'S OFFICE 2023:15

Working paper: Cognitive intrusion

Why hybrid influencing works?

Finnish Behavioural Policy Team (FINBEPOL), Prime Minister's Office



VALTIONEUVOSTON KANSLIA
STATSRÅDETS KANSLI

Publications of the Prime Minister's Office 2023:15

Working paper: Cognitive intrusion

Why hybrid influencing works?

Finnish Behavioural Policy Team (FINBEPOL), Prime Minister's Office

Prime Minister's Office Helsinki 2023

Julkaisujen jakelu

Distribution av publikationer

**Valtioneuvoston
julkaisuarkisto Valto**

Publikations-
arkivet Valto

julkaisut.valtioneuvosto.fi

Julkaisumyynti

Beställningar av publikationer

**Valtioneuvoston
verkkokirjakauppa**

Statsrådets
nätbokhandel

vnjulkaisumyynti.fi

Prime Minister's Office

This publication is copyrighted. You may download, display and print it for Your own personal use. Commercial use is prohibited.

ISBN pdf: 978-952-383-459-0

ISSN pdf: 2490-1164

Layout: Government Administration Department, Publications

Helsinki 2023 Finland

Working paper: Cognitive intrusion Why hybrid influencing works?

Publications of the Prime Minister's Office 2023:15

Publisher Prime Minister's Office

Author(s) Matti Heino, Emmi Korkalainen, Maarit Lassander, Ville Ojanen, Jarno Tuominen
Language English **Pages** 26

Abstract

Hybrid threat campaigns are a spectrum of actions conducted by state or non-state actors, whose goal is to undermine liberal democracies' decision-making, compromise their integrity and saturate their situational awareness and response capacities. Hybrid threat campaigns combine overt and covert military and non-military means. This publication focuses on information influence activities and psychological influencing as forms of hybrid operations.

The effectiveness of hybrid threat campaigns (including actions on the information domain and attempts at social engineering) depends on individual and collective psychological factors and different circumstantial factors. Understanding the psychology of hybrid threats is crucial to better deter campaigns and to mitigate their impact. In this discussion paper, we outline some key psychological concepts and mechanisms of action under the umbrella term cognitive intrusion. This working paper refers to the conceptual model of Hybrid threats developed by Hybrid Coe and the EC's JRC (The Landscape of Hybrid Threats: A conceptual model).

The aim of this working paper is to better understand the extent to which cognitive intrusion mechanisms can increase the effectiveness of hybrid activity and to illustrate how and why they can affect individuals and societies. The purpose of the discussion paper is also to stimulate discussion on how hybrid influence could be combated more effectively by understanding psychological background motives. The aim of the publication is not to provide an exhaustive description of the significance of psychological models and event chains behind hybrid influencing. Here we specifically focus on the underlying psychological mechanisms by which the decision-making capability of individuals and communities can get hampered.

Provision This publication has been carried out as part of the Finnish Behavioural Policy Team activities at the Prime Minister's Office. The content of the publication is the responsibility of the producers of information and the content may not represent the view of the Government.

Keywords Behavioural psychology, Behavioural sciences, Hybrid influencing, Hybrid threats

ISBN PDF 978-952-383-459-0 **ISSN PDF** 2490-1164

URN address <https://urn.fi/URN:ISBN:978-952-383-459-0>

Keskustelupaperi: Kognitiivinen intruusio Miksi hybridivaikuttaminen toimii?

Valtioneuvoston kanslian julkaisuja 2023:15

Julkaisija Käyttätymistieteellinen ennakointi ja tieto tulevaisuuden hallinnossa, Valtioneuvoston kanslia

Tekijä/t Matti Heino, Emmi Korkalainen, Maarit Lassander, Ville Ojanen, Jarno Tuominen
Kieli englanti **Sivumäärä** 26

Tiivistelmä

Hybridivaikuttamisessa on kyse valtiollisten tai ei-valtiollisten toimijoiden toiminnasta, jonka tavoitteena on heikentää liberaalien demokratioiden päätöksentekokykyä, vaarantaa niiden koskemattomuutta sekä kuormittaa niiden tilannetietoisuutta ja toimintavalmiuksia. Hybridi-vaikuttamisen keskeisenä tavoitteena on vaikuttaa kohteen päätöksentekokykyyn. Tässä julkaisussa keskitytään erityisesti informaatiovaikuttamiseen ja psykologiseen vaikuttamiseen hybridivaikuttamisen muotoina.

Tässä julkaisussa hahmotellaan poikkeuksellista termiä kognitiivinen intruusio (cognitive intrusion) korostamaan sellaisia psykologisia ja sosiaalisia painostus- ja tehokeinoja, joilla voi olla vaikutusta hybridivaikuttamisen tehokkuuteen. Julkaisun tavoitteena on havainnollistaa sellaisia tehokeinoja, joilla pyritään aggressiivisesti muuttamaan kohdehenkilön tai -ryhmän tietoisuutta ja käyttäytymismalleja. Tämä voi tapahtua muun muassa ennakkoluuloja, yhteiskunnallisia narratiiveja sekä vakiintuneita asenteita ja uskomuksia hyödyntämällä. Keskustelupaperi etenee yksilötason vaikuttamisesta yhteisötason efekteihin, jotka osaltaan edistävät tai hidastavat yhteiskuntatason ilmiöitä.

Tämän keskustelunavauksen tavoitteena on pyrkiä ymmärtämään paremmin, missä määrin kognitiivinen intruusio voi lisätä hybridivaikuttamisen tehokkuutta, sekä valaista, millaisten mekanismien kautta se voi vaikuttaa yksilöihin ja yhteiskuntiin. Keskustelupaperin tarkoituksena on herättää keskustelua myös siitä, millä tavoin hybridivaikuttamista voitaisiin torjua tehokkaammin psykologisia taustavaikuttimia ymmärtämällä.

Julkaisu pohjautuu Euroopan Hybridiuhkien torjunnan osaamiskeskuksen ja Euroopan komission yhteisen tutkimuskeskuksen hybridiuhkien konseptimalliin sekä aiempiin psykologian alan ja kompleksisuustieteiden käsitteenmäärittelyihin. Tämän keskustelunavauksen tavoitteena on tuottaa ihmiskeskeisempi, käyttäytymistieteellinen näkökulma hybridiuhkien tarkasteluun. Julkaisun tavoitteena ei ole antaa tyhjentävää kuvausta psykologisten toimintamallien ja tapahtumaketjujen merkityksistä hybridivaikuttamisen taustalla.

Klausuuli Tämä julkaisu on toteutettu osana valtioneuvoston kanslian Käyttätymistieteellinen ennakointi ja tieto tulevaisuuden hallinnossa -toimintaa. Julkaisun sisällöstä vastaavat tiedon tuottajat, eikä tekstisisältö välttämättä edusta valtioneuvoston näkemystä.

Asiasanat käyttäytymisen psykologia, käyttäytymistieteet, hybridivaikuttaminen, hybridiuhat

ISBN PDF 978-952-383-459-0

ISSN PDF 2490-1164

Julkaisun osoite <https://urn.fi/URN:ISBN:978-952-383-459-0>

Diskussionsunderlaget: Kognitivt intrång Varför fungerar hybridpåverkan?

Statsrådets kanslis publikationer 2023:15

Utgivare Teamet för beteendepolitik i Finland (FINBEPOL), Statsrådets kansli

Författare Matti Heino, Emmi Korkalainen, Maarit Lassander, Ville Ojanen, Jarno Tuominen

Språk engelska

Sidantal 26

Referat

Hybridpåverkan går ut på att statliga eller ickestatliga aktörer agerar med målet att försvaga liberala demokratiers beslutsförmåga, äventyra deras integritet samt belasta deras lägesuppfattning och handlingsberedskap. Det centrala målet för hybridpåverkan är att påverka objektets förmåga att fatta beslut.

I den här publikationen behandlas en exceptionell term, kognitivt intrång (eng. cognitive intrusion), för att betona olika psykologiska och sociala påtrycknings- och effektmedel som kan inverka på hur effektiv hybridpåverkan är. Målet med publikationen är att belysa olika effektmedel som används på ett aggressivt sätt med avsikten att ändra en målpersons eller en målgrupps medvetenhet och beteendemönster. Detta kan åstadkommas bland annat genom att använda fördomar, samhällliga narrativ samt etablerade attityder och föreställningar. Diskussionsunderlaget behandlar påverkan på individnivå och effekter på samfundsnivå som främjar eller fördröjer olika fenomen på samhällsnivå.

Målet med diskussionsunderlaget är att eftersträva en bättre förståelse av i vilken mån kognitivt intrång kan göra hybridpåverkan effektivare samt belysa genom vilka mekanismer det kan påverka individer och samhällen. Avsikten med diskussionsunderlaget är även att inleda en diskussion om hur hybridpåverkan kan bekämpas på ett effektivare sätt genom att förstå de psykologiska bakgrundsfaktorerna.

Publikationen är baserad på den konceptmodell för hybridhot som Europeiska kompetenscentret för motverkande av hybridhot och Europeiska kommissionens gemensamma forskningscentrum använder samt på tidigare begreppsdefinitioner inom psykologi och komplexitetsvetenskaper. Syftet med diskussionsunderlaget är att skapa ett mera människocentrerat och beteendevetenskapligt perspektiv på granskningen av hybridhot. Syftet med publikationen är inte att ge en uttömmande beskrivning av vilken betydelse olika psykologiska verksamhetsmodeller och händelseförlopp har mot bakgrund av hybridpåverkan.

Klausul Denna publikation har utarbetats som en del av verksamheten för beteendevetenskaplig framsyn och kunskap i framtidens förvaltning. Verksamheten finns i anslutning till statsrådets kansli. De som producerar informationen ansvarar för innehållet i publikationen. Textinnehållet återspeglar inte nödvändigtvis statsrådets ståndpunkt.

Nyckelord Beteendevetenskaper, Beteendepsykologi, Hybridpåverkan, Hybridhot

ISBN PDF 978-952-383-459-0

ISSN PDF 2490-1164

URN-adress <https://urn.fi/URN:ISBN:978-952-383-459-0>

Contents

1	Introduction	7
2	Individual: Vulnerability to cognitive intrusion – Uncertainty reduction as a fundamental principle	10
3	Community: How behaviours spread in communities – Social identity and simple versus complex contagions	13
4	Society: The disruption of the social fabric	17
	4.1 Destabilisation as a shift in attractor landscape	17
	4.2 Societal polarisation.....	21
5	Discussion	23
6	Conclusion	25
	Bibliography for further reading	26

1 Introduction

Hybrid threat campaigns are a spectrum of actions conducted by state or non-state actors, whose goal is to undermine liberal democracies' decision-making, compromise their integrity and saturate their situational awareness and response capacities. Hybrid threat campaigns combine overt and covert military and non-military means.

The effectiveness of hybrid threat campaigns (including actions on the information domain and attempts at social engineering) depends partly on individual and collective psychological factors and different circumstantial factors. Understanding the psychology of hybrid threats is crucial to better deter campaigns and to mitigate their impact. In this discussion paper, we outline some key psychological concepts and mechanisms of action under the umbrella term **cognitive intrusion**.

This working paper refers to the conceptual model of Hybrid threats developed by Hybrid Coe and the EC's JRC (The Landscape of Hybrid Threats: A conceptual model). The model outlines four main pillars of hybrid threats: Actors, Tools applied by the actor, Domains that are targeted, and Phases of activity with the goal to undermine individual and political decision-making capability.

The aim of this working paper is to better understand the extent to which cognitive intrusion mechanisms can increase the effectiveness of hybrid activity and to illustrate how and why they can affect individuals and societies. The purpose of the discussion paper is also to stimulate discussion on how hybrid influence could be combated more effectively by understanding psychological background motives. The aim of the publication is not to provide an exhaustive description of the significance of psychological models and event chains behind hybrid influencing.

Here we specifically focus on the underlying psychological mechanisms by which the decision-making capability of individuals and communities can get hampered. We use the term cognitive intrusion to emphasise the psychological and social levers and effects that condition the impact of hybrid threat campaigns. In this paper, the use of the term *cognitive* refers to a comprehensive understanding of the decision-making parameters. Cognition encompasses reason, emotion, experience, and social situation. Thus, we broaden the use of the term "cognitive" to cover emotional and social effects.

Cognitive intrusion can be seen as the underlying effective mechanism of hybrid influencing. It is a comprehensive set of different forms, methods and means which leverage dispositional and situational factors such as beliefs, prejudices, established political attitudes to impact the behaviour of humans and communities, to shape their value systems, perceptions, and world views. Cognitive intrusion can be seen as the common denominator of the hybrid threat activities in all domains and spheres and provide the lens through which these operations could be perceived. Hybrid threat campaigns can be considered as manoeuvres relying on cognitive intrusion to be effective. This publication focuses on information influence activities and psychological influencing as forms of hybrid operations.

Cognitive intrusion entails transforming the consciousness and behavioural patterns of the target person or group. This entails the manipulation of a particular society or community of its governing characteristics, views, values, and shared norms towards a direction that will push the target to make a decision desired by the aggressor in a given situation. In lay terms, the goal is to intrude into and confuse minds and thereby the social relations and the core of society. The process of cognitive intrusion is incremental but potentially very effective and in the long term has the potential to destabilize and polarize the socio-political sphere of a society.

By intrusion we denote an aggressive and violent violation of boundaries, sovereignty, and the right for self-determination. It refers to the techniques and tactics ranging from hybrid influencing to psychological warfare with the common aim of intruding into the psychosociology of the target individuals and societies. Such influencing is always detrimental to the target or contrary to its own interests.

On the part of the malicious actor, cognitive intrusion is based on an explorative view of human information processing and social behaviour, exploiting our individual and social vulnerabilities. It is a powerful tool on an individual as well as on the social level, as hybrid campaigns take place in social systems. These systems include individuals, communities, and societies, all of which are interconnected, coevolve, and affect each other in reciprocal ways. Therefore, influencing individuals produces emergent effects in their communities, which contribute to societal level phenomena. For example, a new law set by popular demand, feeds back to constrain or enable the behaviour of communities and individuals. Due to the systemic effects, the exact socio-political consequences and effects need not to be deliberated by the actor as the effects are systemic and unpredictable but nevertheless negative, and thus in the interest of the actor. The guiding principle is to sow seeds of doubt, confusion, and mistrust.

The aim of this working paper is to better understand the extent to which cognitive intrusion can make hybrid threat campaigns effective and how and why individuals and societies are affected. We start by recognising the hybrid threat tactics and operations as described and identified in HybridCoe's conceptual model on Hybrid Threats. But we go further by asking why these techniques work, and seek to find answers from the behavioural sciences.

This document is organized based on relevant concepts from psychological and complexity sciences at the levels of individuals, communities, and societies. We describe the human information processing propensities underlying individuals' susceptibility to cognitive intrusion; specifically, uncertainty reduction as a fundamental principle. We take social identity and social contagion (simple versus complex contagions) as the basic explanatory frameworks for the spreading of behaviour in communities. Finally, the disruption of the social fabric of the target society is explained by the complexity science concept of an attractor landscape. We end with key questions and recommendations to map the terrain for future research and action.

2 Individual: Vulnerability to cognitive intrusion – Uncertainty reduction as a fundamental principle

The *predictive processing view*¹ suggests that humans and other animals are intrinsically motivated to optimize their internal models of the external world to avoid surprise, i.e., to reduce uncertainty. This theoretical principle first proposed in the 19th century to organize our understanding of human behaviour, has undergone a renaissance due to its applicability to shed light on various longstanding questions and findings related to the neurosciences and psychology. Our experience is structured around internal working models, or schemas, about ourselves, others, and immediate environment. Uncertainty reduction is the underlying principle governing human information processing and explains our vulnerabilities to cognitive intrusion.

Within the human brain this uncertainty reduction is realized hierarchically from lower-level sensory predictions, emotional states, thoughts and eventually to higher order belief structures about the world and the self. Whenever there is a mismatch between the expectation and the experience, a prediction error signal is generated. This mismatch is then attempted to be removed in a hierarchical fashion. More specifically, there are two ways to deal with the uncertainty: update the internal model, i.e., one's expectations and beliefs — or alternatively, change one's reality to harmonise with the internal prediction. Facing psychologically inconsistent or cognitively dissonant situations, people tend to change the situation until their experience becomes consistent².

A common error in our attempts to counteract malicious influencing is to ignore the cognitive, social and affective aspects of uncertainty reduction. In the case of mis- and disinformation, public counter measures often focus on access to facts, whereby a thorough explanation of facts should overcome the impact of mis- and disinformation. However, research shows that the rejection of e.g. climate science is not the result of

1 Wheeler N. E., Allidina S., Long E. U., Schneider S. P., Haas I. J., Cunningham W. A. (2020), Ideology and predictive processing: Coordination, bias, and polarization in socially constrained error minimization

2 Festinger L. (1957) A Theory of Cognitive Dissonance.

ignorance, but is affected by conspiratorial mentality, fears, identity expression and reasoning. It is driven by the need to reduce uncertainty between personal values and the interpretation of the situation, rather than objective evidence.

Uncertainty reduction has many well documented consequences on human information processing known as cognitive biases. For example, when deciding what is true, people are often biased to believe in the face validity of information and 'go with their gut' and intuitions instead of deliberating. For this effect, simply repeating any claim- true or false- makes it more believable than presenting it only once. This **illusory truth** effect arises because people use **cues such as familiarity** (a message has been encountered before), processing fluency (a message is effortlessly understood) and **cohesion** (a message integrates with existing memory structures) as evidence for truth, and the strength of these cues increases with repetition.

Furthermore, instead of active deliberation, we are prone **to intuitive and superficial thinking** as well as defaulting to our own existing personal views. Instead of questioning and reconsidering, we focus on information supporting our worldview. It should be noted that this is not a pathology; in uncertain environments, **relying on heuristics or "rules-of-thumb"** can be a far superior strategy, compared to extensively deliberating on every minute and transitory detail available³. Ignoring information inconsistent with our existing worldviews, also serves us when faced with disinformation that aims to challenge our perceptions. Practically useful heuristics often emerge from the grass-root level; they can be studied -and their dissemination coordinated by agents interested in promoting community resilience or psychological defence. Heuristics become a problem when successfully hijacked by malicious actors, in which case cognitive structures based on misinformation can be hard to dismantle (even producing false memories and facts). This can be seen in a lingering influence of misinformation on people's reasoning, after it has been corrected – an effect known as the **continued influence effect**.

Emotions play a crucial role in human information processing. For example, appeals to emotion increase persuasion and generating fear can successfully change attitudes, intentions and behaviours. Research has suggested that sad mood might reduce gullibility, anger has been shown to promote belief in politically concordant misinformation, and negative mood from social exclusion can increase susceptibility to conspiratorial content.

3 Hogarth, R. M. (2012). When Simple Is Hard to Accept. In *Ecological Rationality: Intelligence in the World* (pp. 61–79). Oxford University Press.

As cognitive functioning benefits from experiencing coherence, instilling uncertainty to diminish this coherence becomes key in destabilising an opponent. A malicious actor would subsequently push alternative higher order explanations, that are tailored to reduce uncertainty and simultaneously target one's **basic psychological needs** (i.e., the need for belongingness, autonomy, and competence)⁴. A malicious actor aims to instil uncertainty overall, for example, especially by targeting the social referents of truths: that is the structures that create a sense of shared reality, such as the media, government, or communities. To reduce this uncertainty, we're liable to adopt more simplistic or black-and-white beliefs than we normally would.

For example, adopting a belief that a local governance is corrupt will lead to uncertainty in accompanying realms, such as voting behaviour, taxation, and the agency of the individual in their social surroundings. At worst, this can lead to a cascading effect, where one attempt to reduce uncertainty will result in major changes in the belief landscape overall, possibly polarizing them further.

The uncertainty reduction principle provides a way to incorporate human cognitive, emotional, and social processes under the same umbrella mechanism. Of special relevance here, is that as humans we often use others as means to reduce uncertainty. This places social networks and communication as integral in how we change our minds and what we believe in.

4 Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

3 Community: How behaviours spread in communities – Social identity and simple versus complex contagions

Any group can be understood as a social network; individuals connected by strong, weak or no relationships. Behaviours spread through these links, as social connections and networks heavily influence the factors underlying behaviours – such as those pertaining to capabilities, physical and social opportunities, emotions, motivations, and basic psychological needs. Human connections and relationships are understood as a core need, and the desire to connect as a fundamental drive, ultimately stemming from the social functions of the human brain. Individuals do not form their mindsets in isolation, but they are adopted from the surrounding social realms and integrated into a functional internal working model.

The social behaviour of humans is based on our fundamentally social brain. The social brain theory posits that the main driver in primate brain evolution has been the increase in social complexity, which has placed a demand for cognitive resources to better coordinate, compete and keep track of the dynamic social ties and gossip⁵. The social brain builds a picture of the human brain as a social organ that, through its functions, can set the basic parameters of how we relate to each other and behave like we do, depending on the nature of these relationships.

Social identity refers to the ways that people's self-concepts are based on their membership in social groups and their attitudes towards other groups. By categorizing oneself as a group member, one acknowledges possession of the group-defining characteristics shared by others who are similarly categorized. The degree of social sharedness is the degree to which cognitions, preferences, identities, etc. are shared within groups. A group identity is built upon psychological ingroup-outgroup -dynamics. An in-group is a social group to which a person identifies as being a member, and contrarily an out-group is a social group which one does not identify with.

5 Dunbar R. I. (2009). The social brain hypothesis and its implications for social evolution. *Annals of human biology*, 36(5), 562–572.

The increase in polarized political views and group identities is an extreme example of the spreading of ideas and behaviours in a social network. In the last decade, the distance between political parties and the irreconcilable differences (either ideological, personalistic, or both) among voters has increased worldwide⁶. Over the past decade, polarisation has been both a key tool and a target for information-psychological influencing. Influencing seems to be increasingly linked to the attitudinal atmosphere of liberal democracies (such as attitudes towards minorities, gender roles, energy policy etc.) and thus to people's everyday experiences and ideas. Accelerating socio-political polarisation is in the interest of malicious actors, as it weakens the essence of liberal democracy. It creates division and may lead to crises, disagreements that would not find solutions in the regular channels of democratic debates. It might also lead to more violent political expression, brutalization of political interaction, having the potential to damage the institutions essential to democracy. The amount of democracy has been deteriorating globally over the past 15 years^{7 8 9 10}, which is for one reflected in the political and ideological polarisation of the information environment. From a broad perspective, the development of polarisation includes the risk of increasing tensions and social unrest.¹¹

Social contagion¹² refers to the ways ideas and behaviours spread in social networks, – both online and offline. One might argue that the optimal way to “infect” the network with novel norms would be to target highly connected individuals with few overlapping connections. Recent work on the subject indicates that this is indeed how information that is already accepted by the receivers – especially within ingroups – tends to spread. In e.g. social media, people share widely and non-critically the kinds of posts that reinforce their existing beliefs. One contact can be enough to “infect” the person with a novel idea that fits within the existing network of beliefs. This process is termed **simple contagion**.

There are many mechanisms and consequences of simple contagion, i.e. behaviours spreading within the echo chambers of likeminded peers, starting from the fact that people believe in-group members more than out-group members, and tend to view ingroup members more favorably than those of an out-group. In-group messages

6 Reiljan, 2020.

7 Global State of Democracy Initiative: Global State of Democracy Report 2022, Forging Social Contracts in a Time of Discontent

8 Economist Intelligence: Democracy Index 2022

9 V-Dem Institute: Democracy Report 2023, Defiance in the Face of Autocratization

10 Freedom House: Freedom in the world 2023, Marking 50 Years in the Struggle for Democracy

11 Government Report on Internal Security, Finnish government 2021:48, pp. 12.

12 Centola, D. (2021). Complex contagions. In Research Handbook on Analytical Sociology (1st ed., pp. 321–335). Edward Elgar Publishing.

are more persuasive and seem more true, as they come from sources perceived to be credible rather than non-credible. People trust human information sources more if they perceive the source as attractive, powerful, and like themselves. In the field of information influencing, this may be one of the most effective mechanisms of influencing, as it is linked to trust in a way that may be difficult to undermine from the outside.

Putting undue emphasis of voiced opinions on the expense of those left unvoiced, can lead to the perception of a consensus where there is none. This can be exploited by trolls and bots, when people become increasingly hesitant to voice views discordant with the apparent but false group consensus. We often overlook, ignore, forget, or confuse cues about the source of information, which explains why a small number of social media accounts can succeed in spreading an outsized amount of misleading content. Experts and political elites are trusted by many and have the power to shape public perceptions; therefore, it can be especially damaging when leaders make false claims – also increasing their appeal as targets of distorted information. People tend to share information which triggers emotional response and may lead to sharing more extreme content in general. Things shared to a greater degree have stronger effects on group behaviour than those shared in lesser degree.

Like-minded ideas that strengthen ingroup-outgroup contrast and dislike for the opposing party effectively amplify polarisation¹³. As social identity is heightened, depersonalization occurs and nongroup relevant aspects of the self are diminished. Simultaneously positive aspects of the group support self-worth. Even unethical behaviour can be morally justified within the in-group as far as it serves the interest of the group.

In conclusion, simple contagion refers to a situation where the information is already accepted by the receivers, and hence merits only superficial processing before sharing. However, we can also identify a process of complex contagion: ideas and behaviours traveling differently through the network, when they are deemed socially “risky”. This includes cases where the ideas and behaviours challenge the existing norms, require social coordination or emotional excitement to be useful, or are not clear-cut regarding their legitimacy. For a person to take up such a novel behaviour, they need to observe several people around them adopting the behaviour first. In this case, transmitting the message within e.g. a community requires a different strategy than that of

13 See: Affective polarization: Yengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual review of political science*, 22, 129–146.

non-overlapping “influencers” in the case of simple contagion: multiple advocates, whose connections overlap significantly, need to be recruited so that community members can observe more than one person taking up the behaviour or spreading the idea.

The process of adopting polarized political views and group identities is an example of complex contagion. Adopting radical or extremist group identity becomes more tempting for people who experience uncertainty regarding their place in the world, as well as what the future holds. Strong social sharedness and group identity then decreases the felt uncertainty, especially when the uncertainty is a threat to one’s identity.

The type of groups that most effectively reduce identity-related uncertainty are those that are highly entitative. These groups give the individual rigid and closed boundaries, clear ideological belief systems, a simple black and white worldview, and often a charismatic and authoritarian leader. Individuals high in uncertainty gravitate towards groups high in entitativity, as means of restoring certainty regarding the self.

As social polarisation advances, the general uncertainty within the community increases, and disrupts its functioning. The society destabilizes and begins to function in a segregated fashion at the expense of integration.

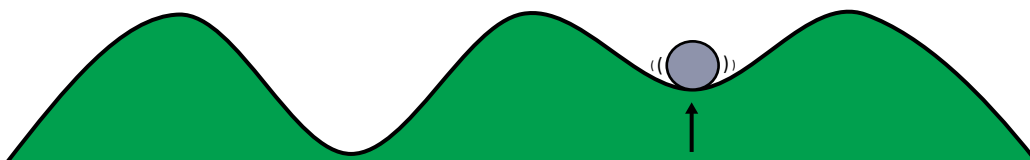
4 Society: The disruption of the social fabric

4.1 Destabilisation as a shift in attractor landscape

Systems can change their state in two broad ways – abruptly, or gradually – and important societal changes often happen through relatively abrupt transitions¹⁴. Attractor landscapes present a conceptual model to understanding abrupt change in both physical and social systems, including but not limited to individuals, communities and societies.¹⁵

The basic idea is to consider the system to have several states at its disposal; these are visualised as valleys in figure A below, while the ball represents the system's current state. The shape of the landscape is defined by the capabilities, motivations, and opportunities available to the system.

Figure 1. Figure A



Let us consider an example (figure A) where the attractor (valley) on the right represents a baseline state “members of the society have strong trust in governmental institutions”; and the attractor on the left represents a state “members of the society distrust governmental institutions”. The shape of the landscape is determined by the dynamics of social influences (e.g. considering others perceive the institutions as trustworthy), physical

14 See <https://psyarxiv.com/3rxyd> for an entry-level introduction in the behaviour change context

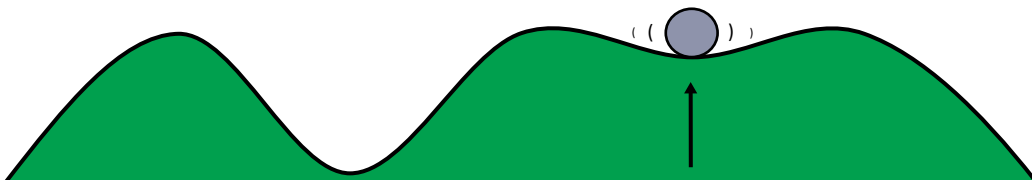
15 Heino, M. T. J., Proverbio, D., Marchand, G., Resnicow, K., & Hankonen, N. (2022). Attractor landscapes: A unifying conceptual model for understanding behaviour change across scales of observation. *Health Psychology Review*, 1–26.

opportunities (e.g. ability to participate in decision making), motivational factors (e.g. social identity), capabilities (e.g. ability to distinguish misinformation) and behaviours (e.g. information sharing) of the population.

A hybrid threat actor attempts to erode trust between the state institutions and the people as well as between various populations. A key objective is to increase uncertainty disrupting the relationship within and between different populations and the government resulting in deepening social cleavages and weakening social cohesion. Malicious actors provoke internal divisions through a variety of tactics, including the “4 Ds of disinformation” – dismissing, distorting, distracting, and dismaying¹⁶. Contemporary digital and social media platforms allow malicious actors to influence these relations with considerable ease.

Successful hybrid campaigns can destabilise the current state (figure B) by e.g. changing **societal narratives** that pertain to trust in institutions, leading to increased temporal sway in relevant indicators of institutional trust.

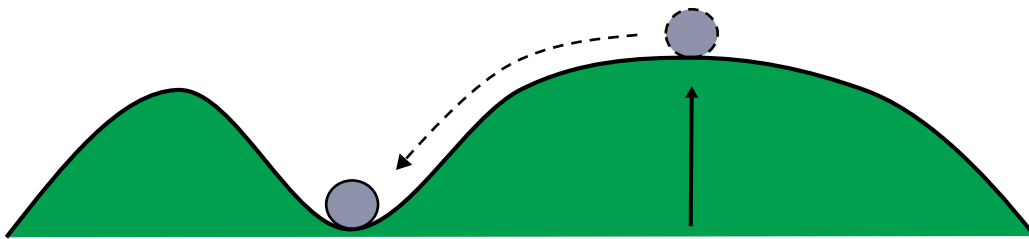
Figure 2. Figure B



16 Nimmo B. (2018), e.g. Combatting disinformation with the four D's, Center for Academic Innovation, University of Michigan

Finally, when destabilisation efforts are complete (figure C), the system is amenable to any small push, which will cause it to change its state.

Figure 3. Figure C



Another way for a system to change its state abruptly is, if the landscape remains constant, but some event or intervention pushes the system past a tipping point (consider the ball rolling over the hill to the leftmost attractor in (Figure A). In this case, the earlier state remains accessible; unlike in the case where changes in the landscape itself were the primary catalyst for the transition (Figure C). Abrupt changes in real systems take place as some combination of these two categories, and the number of possible attractors is not limited to two.^{17 18 19}

The above presented mechanisms and concepts underscore the vulnerability of individuals and groups for external malicious actions. The systemic view gives rise to an understanding of what happens when the disruption is successful: National or supranational cooperation is disrupted, and the system increases in uncertainty. This may lead to reaching a “tipping point” (depicted by hilltops in the figure above, Figures A-C), resulting in an abrupt state change within the organization.

This can happen, when the system’s normal functioning is challenged to the extent, where upon release of the stress to the system, it no longer returns to its original state, but wanders off to a new attractor. Tipping phenomena are difficult to predict as they are by nature nonlinear and dependent on the initial state of the system. In retrospect, it is often

-
- 17 Scheffer M., Carpenter S., Lenton T., Bascompte J., & Vandermeer J. (2012). Anticipating Critical Transitions. *Science magazine*
 - 18 Lever J., van de Leemput I., Weinans W., Quax R., Dakos V., van Nes E., Bascompte J., & Scheffer M. (2019) *Foreseeing the future of mutualistic communities beyond collapse*, Wiley Online Library.
 - 19 Scheffer M., Borsboom D., Nieuwenhuis S., & Westley F. (2022). Belief traps: Tackling the inertia of harmful beliefs, *Proceedings of the National Academy of Sciences*.

possible to observe, that the moment of crossing a tipping point was a *critical juncture*²⁰ in the development of the system's path, even though viewed prospectively, the event was a so-called black swan²¹ to all or most contemporary actors.

Such tipping can take place on multiple scales, in a cascading fashion: For example, disinformation in social media causes an individual to increasingly doubt their possibilities to influence their environment, until they shift from the attractor "an active participant" in local democracy and decision making, to one of "a passive observer". This change may affect the state of the larger community. For example, the number of people actively taking part in local decision-making drops, until the community switches from an attractor state "active and able to self-organise and adapt" to one of "passive and / or overly polarised to take initiative".

The theory underlying attractor landscapes points to two important issues in risk management. Firstly, large changes need not be the result of large events, but small pushes can suffice, when the system resides in a shallow attractor or on the top of a "hill" in the landscape. Secondly, the fact that earlier events (e.g. misinformation campaigns of a specific type) have not produced adverse consequences, does not imply that they continue not do so in the future. In other words, when the underlying landscape has shifted to being receptive to change (the current attractor has become shallow), formerly ineffective campaigns can gain potency.

An interesting question is, if there is a way to keep track of e.g. societal or community attitudes and trust in a way that allows one to observe and act upon worrying destabilisation trends before the current, desirable state reaches a critical point. The increase in sway when an attractor becomes shallower is of relevance, when assessing the possible futures of a system. With dense (e.g. daily) data, it is possible to monitor indicators for early warning signals of destabilisation, which has the potential for pushing a system past a tipping point. Such data has been collected in the context of the COVID-19 pandemic via social media²².

20 Soifer H., (2012) *The Causal Logic of Critical Junctures*, Comparative Political Studies, Sage.

21 Taleb, N. N., & Blyth, M. (2011). *The Black Swan of Cairo*. Foreign Affairs.

22 Astley, C., Tuli G., Mc Cord K & Brownstein J. (2021). Global monitoring of the impact of the COVID-19 pandemic through online surveys sampled from the Facebook user base. *Proceedings of the National Academy of Sciences, Biophysics and computational biology*.

4.2 Societal polarisation

Hybrid threat campaigns benefit from mistrust between people and political communities, which means that efforts can be made to strengthen distrust by various means. In Western Europe and the United States, trust levels have been steadily declining since the 1970s and state institutions are losing their credibility owing to diminishing public trust²³.

In some circumstances a government can also become increasingly distrustful of their citizens. This can be due to e.g. a small number of people with strong polarised opinions who have gained a large share of the societal voice, as individuals with more nuanced views have become passive, leading to the aforementioned false consensus effect. As communities have also lost much of their ability to self-organise, the implicit governance strategy switches from “provide freedom to make many local decisions” to “command and control”. The switch is then capitalised by hybrid influencers (e.g. narratives such as *“the government wants to control you, and you should reciprocate their distrust”*) to radicalise people on the individual and community levels. A key psychological concept underlying our relationship to socially transmitted information, and thus also to its misuse, is **epistemic trust**. It denotes a multifaceted process to consider the trustworthiness, relevance, and quality of both the information and the informant. In short, epistemic trust can be divided into three dispositions of orienting toward other people: we may either mistrust others as credible sources of information; to trust people as credible sources of useful knowledge or information while appropriately evaluating and updating their credibility; or to overly trust (i.e., credulity) others even when there is little basis for this trust. It is especially this last group who are most susceptible for dis- and misinformation.²⁴
²⁵ Furthermore, it is known that conspiracy mentality lowers trust in authoritative information sources, yet is correlated with a higher trust in the perceived credibility of lay sources²⁶.

Also, people’s trust in each other remains an equally important target of influencing. The rise of populism is built on the ease of sowing mistrust between people. Populist communicators rely on persuasive strategies by which certain social group cues are made to affect people’s judgment of and political engagement with political issues. The populist

23 United Nations, Department of Economic and Social Affairs: Trust in public institutions: Trends and implications for economic security, 2021

24 1. preprint: Tanzer M, Campbell C, Saunders R, et al. Acquiring knowledge: Epistemic trust in the age of fake news. PsyArXiv; 2021. DOI: 10.31234/osf.io/g2b6k.

25 Nimbi, F. M., Giovanardi, G., Baiocco, R., Tanzilli, A., & Linguardi, V. (2023). Monkeypox: New epidemic or fake news? Study of psychological and social factors associated with fake news attitudes of monkeypox in Italy. *Frontiers in Psychology*, 14.

26 Imhoff, R., Lamberty, P., & Klein, O. (2018). Using power as a negative cue: How conspiracy mentality affects epistemic trust in sources of historical knowledge. *Personality and Social Psychology Bulletin*, 44(9), 1364–1379.

society is separated into antagonistic groups, a strategy called 'populist identity framing': The ordinary people "us" are portrayed as being threatened by various out-groups "them" and while the political elite benefits from having the upper hand, the people are threatened. Turning away from reliable mainstream media to "filter bubbles" and echo chambers further confirms one's ingroup identity and beliefs. In addition, objective truth is not considered important as information is considered true, only as so far as it benefits the ingroup.

Also in recent years, the level of political and social authoritarianism and the percentage of voters with populist authoritarian views in Western countries has become a concerning and significant issue.

Theories of authoritarianism share two fundamental ideas: (a) authoritarianism is mainly an individual response to threatening situations, and (b) although there may be a certain predisposition to authoritarianism in some people, contexts that threaten security and social order can favor authoritarian attitudes in any individual through the experience of not having the cognitive, emotional, and social resources to deal with the situation.²⁷

Authoritarianism gets its power from the individuals' perceptions of anomie, that is, the *view of the world as an essentially insecure and dangerous place*. Threat of economic crisis and low socio-economic status are correlated with a view of the world as excessively dangerous. When *individuals perceive a lack of control over their environment, they are more likely to adopt ideologies that allow them to regain the feeling that their environment is structured and predictable and increase perceived control*.^{28 29}

Polarized societies are especially vulnerable to divisive and fear-inducing populist messages from authoritarian leaders- "life as we know it" is under attack. The ease with which authoritarian leaders influence their followers is partly explained by the dispositional characteristics of the individuals drawn to authoritarian leaders and the way the individuals are transformed by the leaders and their ingroups. This development might have the potential to lead to radical tipping points through accelerationism: situation is so dire, that the *only option is to completely dismantle society's governing ideologies in the aim of replacing them with better ones*³⁰.

27 Torres-Vega L., Ruiz J, Moya M. (2021). Dangerous Worldview and perceived sociopolitical control: Two mechanisms to understand trust in authoritarian political leaders in economically threatening contexts, *Frontiers of Psychology*.

28 Torres-Vega L., Ruiz J, Moya M. (2021). Dangerous Worldview and perceived sociopolitical control: Two mechanisms to understand trust in authoritarian political leaders in economically threatening contexts, *Frontiers of Psychology*.

29 Sibley C. G., Duckitt J. (2013). The dual process model of ideology and prejudice: a longitudinal test during a global recession. *J. Soc. Psychol.* 153 448–466.

30 Smith C, Tindale S. A Social Sharedness interpretation of the January 6th U.S. capitol insurrection, (2022). *Group dynamics theory reaserch and practice*.

5 Discussion

Nations generally aim to maintain a predictable organization, where international agreements and cooperation and organized means of competition allow uncertainty to remain tolerable. Within democratic states the legitimacy stems from the functioning of democratic institutions, which support the individuals' psychological needs of agency, competency and belongingness. This is why democracies are especially vulnerable to attacks on these institutions, and why they seem primary targets for disruption. Decreased citizen participation, trust and belief in justness of the nation all destabilize the system and disrupt its functioning.

Collective resilience refers to how shared identification within a group allows its members to deal with adversity. This takes place via a process of enacting and expecting solidarity and cohesion, which enables coordination and drawing upon collective resources.³¹

In their list of 12 actionable recommendations to facilitate collective psychosocial resilience with the public, Drury et al. point to the importance of working with, not against people's group norms and social identities.³² The key to doing this a) knowing what the norms and identities are, and b) making sure there is at least *potential* for a shared identity with emergency planners and responders. Societies with a history of high institutional trust, tend to have generated efficient procedures of one-way communication, as this has sufficed when working with a relatively homogenous public.

Polarisation and migration – due to e.g. war, climate change or commodity crises – challenge this homogeneity by instilling novel social groups within countries. This represents a challenge for communication, as it becomes imperative that the authorities engage in a two-way conversation with target groups, and simply mass-tailoring media

31 Drury J, Cocking C, Reicher S. Everyone for themselves? A comparative study of crowd solidarity among emergency survivors. *Br J Soc Psychol.* (2009) 48:487–506.

32 Drury, J., Carter, H., Cocking, C., Ntontis, E., Tekin Guven, S., & Amlôt, R. (2019). Facilitating Collective Psychosocial Resilience in the Public in Emergencies: Twelve Recommendations Based on the Social Identity Approach. *Frontiers in Public Health*, 7, 141.

messages no longer suffices. But this is a huge strain on potentially already overburdened and over-optimised strategic communication frameworks. How should authorities respond to this challenge?

One way to target such influencing is by strengthening one's capacity to tolerate uncertainty by providing means to realistically evaluate the situation and source, and strengthen the social fabric in a way that maintains normal discussions as a form of reality-check. Focusing on the nature and quality of relationships within and across communities is the cornerstone of efforts to neutralise hybrid threats. This requires sustained behavioural scientific capabilities and efforts at the structural and policy levels to develop reciprocal links between the state and the people as well as fostering cohesion within and between communities.

6 Conclusion

In a world of increasing complexity and fast communication it is important to prepare for the perturbations from malevolent actors. Understanding both the psychological and systemic mechanisms by which such influencing generally takes place is necessary to defend against such actions and mitigate against its impacts. Here we have reviewed select phenomena, such as uncertainty reduction, social phenomena, and attractor landscapes, all which envelop the interlocking layers between individuals, collectives and broader systems. While aiming to provide a general review a more detailed discussion is warranted on how to best make use of these phenomena to safeguard against cognitive intrusion. This should include not only basic research on the mechanisms of influence, but also systematic research on ways to intervene on such efforts. The adversarial dynamic between the malicious actors and those targeted calls for a continued evaluation and development of mitigation tactics. Socio-psychological elements of hybrid influencing should also be identified as part of the overall development of comprehensive security and strategic thinking. For this end behavioural sciences have an important contribution to make, and should be included in the relevant discussions.

BIBLIOGRAPHY FOR FURTHER READING

- Chevance, G., Baretta, D., Heino, M. T. J., Perski, O., Olthof, M., Klasnja, P., Hekler, E., & Godino, J. (2021). Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. *PLOS ONE*, 16(5), e0251659.
- Hasselman, F. (2022). Early Warning Signals in Phase Space: Geometric Resilience Loss Indicators From Multiplex Cumulative Recurrence Networks. *Frontiers in Physiology*, 13, 859127.
- Proverbio, D., Kemp, F., Magni, S., & Gonçalves, J. (2022). Performance of early warning signals for disease re-emergence: A case study on COVID-19 data. *PLOS Computational Biology*, 18(3), e1009958.
- Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I. A., Levin, S. A., van Nes, E. H., Pascual, M., & Vandermeer, J. (2012). Anticipating Critical Transitions. *Science*, 338(6105), 344–348.
- Schiepek, G., & Strunk, G. (2010). The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—A method for the real-time monitoring of human change processes. *Biological Cybernetics*, 102(3), 197–207.
- Weinans, E., Lever, J. J., Bathiany, S., Quax, R., Bascompte, J., van Nes, E. H., Scheffer, M., & van de Leemput, I. A. (2019). Finding the direction of lowest resilience in multivariate complex systems. *Journal of The Royal Society Interface*, 16(159), 20190629.



VALTIONEUVOSTON KANSLIA
STATSRÅDETS KANSLI

SNELLMANINKATU 1, HELSINKI
PO BOX 23, 00023 GOVERNMENT, FINLAND
Tel. +358 295 16001
info.vnk@gov.fi
vnk.fi/en

ISBN pdf: 978-952-383-459-0
ISSN pdf: 2490-1164