Building mental resilience against disinformation: an experiential futures case study

by

James S. Law

Submitted to OCAD University in partial fulfillment of the requirements for the degree of a

Master of Design in Strategic Foresight and Innovation

Toronto, Ontario, Canada, 2023

Abstract

Disinformation, its impacts and mitigation measures have been a focus of governments, communities and industry as globalized societies struggle to govern, co-operate and build policy under rapid digitalization. As regulating the spread of disinformation is difficult, this research project explores the emerging futures and interventions supporting resilience against disinformation and how the public receives and reacts to these futures. This project also explores the use of experiential futures as a tool in foresight and prototyping to sample public feedback. Literature and media reviews and expert interviews were used to generate experiential futures depicting extreme polarities of interventions to build resilience against disinformation: high government regulation of the use of tech, commercial protection for individuals against disinformation and more community and education-focused efforts. The futures were physical installations created using a combination of physical items and audio-visual elements using the POEMS framework. Members of the public were invited to navigate these futures and provide their responses, reactions and perspectives using a feedback questionnaire. Results indicate that individuals are most fearful of government and commercial/industry interventions and prefer community-based approaches. This is aligned with preferences and legal limitations on how disinformation generation and spread is regulated. Through the use of experiential futures, participants were able to identify the differences in the futures and shared their different reactions and responses to each future. It is proposed that with technological advances helping make new immersive and creative experiences more affordable and scalable (e.g. VR, AR, 3D printing), there is opportunity to understand whether investing time and resources in making experiential futures more immersive will generate different or more insightful responses and feedback from participants.

Table of Contents

List of Figures

List of Tables

1. Introduction

In recent years, concern over disinformation has increased as its volume, reach and impacts on individuals and communities have materially affected the ability of societies and governments to effectively co-create and implement collective responses to global issues. Most recently, the influences of disinformation can be seen in individual and community responses to the COVID-19 pandemic, where disinformation spread via social networks even convinced people to destroy 5G cell towers in vigilante movements to stop the spread of the virus (Ahmed et al., 2020). There are also examples of less acute disinformation sharing – though with equally substantial impact – such as the longer-standing campaigns to downplay the anthropogenic nature of climate change, contributing to the lack of a consistent and co-ordinated global response (Cook, 2019).

The increased adoption of social media platforms and technology to communicate and share information has enabled disinformation to be spread more easily, widely and quickly (Southwell et al., 2018). While governments have attempted to regulate disinformation by engaging social media and technology companies, they have largely been powerless (Havelin, 2021). Companies like Meta have near-monopolies over how media is shared in the digital age, equipping them with significant power and consumer support that can be leveraged against government interference (Pickard et al., 2020). In situations where there is co-operation from industry, regulations concerning user-generated and shared content are difficult to design and implement. Developing such regulation requires defining what is and is not disinformation, engaging with how this impacts individual freedom of speech and expression (Pielemeier, 2020) and debating whether regulating content is aligned with the principles of a democratic society (Tan & Sijie, 2020).

With disinformation supply difficult to regulate, some researchers have turned their attention to understanding how individuals and communities can become more resilient to disinformation. Scanning the literature, these methods range from general media literacy in the classroom to more direct interventions that aim to "inoculate" against disinformation (Roozenbeek et al., 2022). These latter, more direct inventions may include asking individuals to create and then deconstruct disinformation on highly political topics (e.g. immigration) to issuing direct warnings about upcoming disinformation campaigns (e.g. Russian propaganda about the war in Ukraine). However, like content regulation, these more direct methods can be perceived to be highly personal, interfering with individual thinking, expression and psychology and thus make them difficult to implement. Recent experiences with debates over vaccines introduced to combat the coronavirus and COVID-19 demonstrate the challenges associated asking individuals to take an action with their body in contribution to the greater good.

As a consequence of the above considerations, testing public perception and receptivity to methods of building resilience against disinformation can help policymakers and practitioners understand their likelihood of adoption. In line with design thinking principles, exposing stakeholders to early prototypes or environments of solutions is an important way to gather this feedback on a concept's success, weak points and emerging form. More common prototyping methods like UX testing or storyboarding help researchers understand individual reactions to specific tasks or actions. However, they do not adequately immerse research participants in a more holistic environment that involves testing emotional, political and broader sensory responses – elements that can help researchers gauge responses to policies and social interventions.

To explore more immersive options, this project also examines the use of experiential futures to gauge public reception to various scenarios depicting how individuals and communities may build resilience to disinformation in the future. Experiential futures is one method that may help researchers gain a more holistic understanding of participant reactions to different futures. Falling into the broader category of design fiction, experiential futures enable participants to experience a 'slice of life' of a future scenario through the use of roleplay, audio and visual media, and physical artefacts. By participating in an experiential future, individuals can imagine themselves in the scenario that stretches beyond the near future (Bleecker, 2022), allowing researchers to observe and sample reactions.

To set-up these experiential futures, this project first explores trends around how governments and institutions are building individual and community resilience against the impacts of disinformation, to understand the underlying drivers. The most impactful and uncertain drivers of change will be used to synthesize different future scenarios reflecting how society, governments and communities may help build resilience against disinformation. These scenarios were turned into experiential futures through the creation of artefacts and audio/visual media displayed at the democracyXchange summit. Attendees who experience these scenarios were sampled for their reactions to better understand how communities may receive mental resilience interventions. The reactions also provided insight into the strengths, gaps and opportunities of using experiential futures methodology in assessing public receptivity and responses to different futures.

2. Context

Origins of disinformation and its evolution

Reviewing the literature, false or incorrect information is often categorized by the intention behind its creation and spread. For example, false information is broadly categorized as misinformation – a factually incorrect idea, notion or theory that spreads without any malice or negative intent behind its creation (Fallis, 2014). An example of this is the pop culture emergence of the "Mandela effect", where entire groups of individuals are insistent that they have memories of popular events, book titles or brand logos being different than what is real (Figure 1). The name of this effect was coined by Fiona Broome who, along with others, recalled the death of South African president Nelson Mandela, despite this having never occurred (Prasad & Bainbridge, 2021).
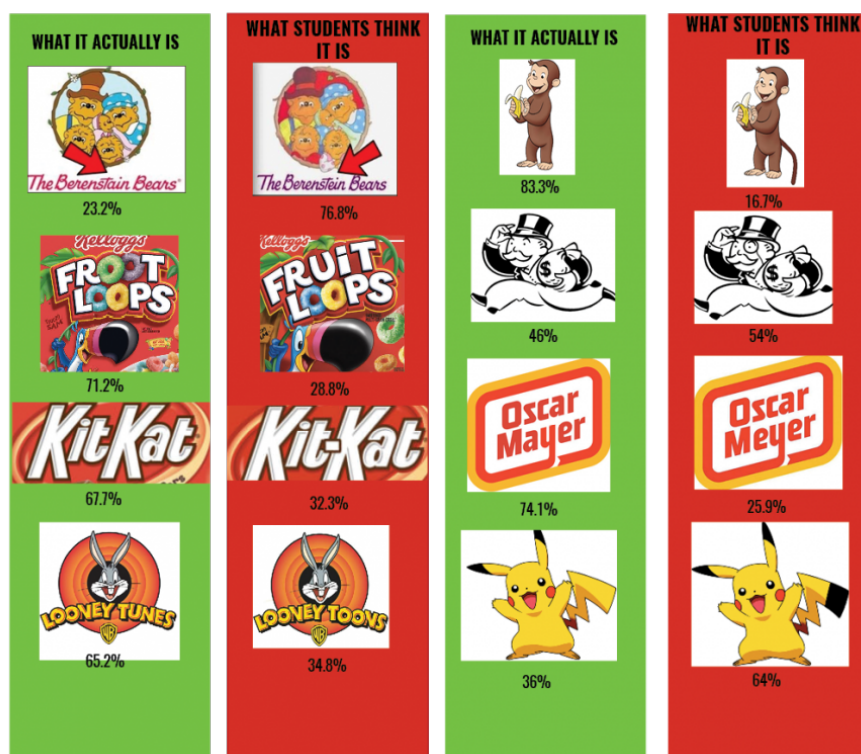
Figure 1- Representing the Mandela Effect - A sample of images that people recall versus the real and original images.

There is also malinformation, a fact that is used in an incorrect context, but used to embellish a story and create an emotional response from the reader or observer – for example, using a photo of an empty grocery store from the mid-2000's to embellish a story about pandemic supply chain issues from 2020 (Wardle, 2022). While it is true that there were supply chain issues, any anecdotes from readers referring to that image specifically, would have been false.

However, disinformation is misinformation that was created and spread with the intent to deceive and create a desired reaction in society – even if that desired reaction is chaos and conflict (Fallis, 2014), contrasting the more benign intents behind general misinformation. For example, while the Mandela Effect is seen as fun and amusing by some, others have used this to incite conspiracy theories that there is a memory manipulation program in effect by the occult (French, 2019) - turning misinformation into disinformation. When this disinformation is spread widely with the intent to cause significant reactions and social volatility, its impact can trigger violent and disruptive actions such as the destruction of infrastructure, refusal of vaccines and preventing co-operation in solving global challenges such as climate change.

While challenges associated with disinformation may seem more recent and acute (i.e., the terms "fake news" and "alternative facts" were popularized in the mid-2010's) (Figure 2), the problems associated with the spread of incorrect information or false information are not new. Notably, the term disinformation emerged in the 1960's during the Cold War, likely originating from the Russian word, "*Dezinformatsiy*a" which referred to the KGB propaganda department that was created to spread false information to impact American morale (Taylor, 2016).
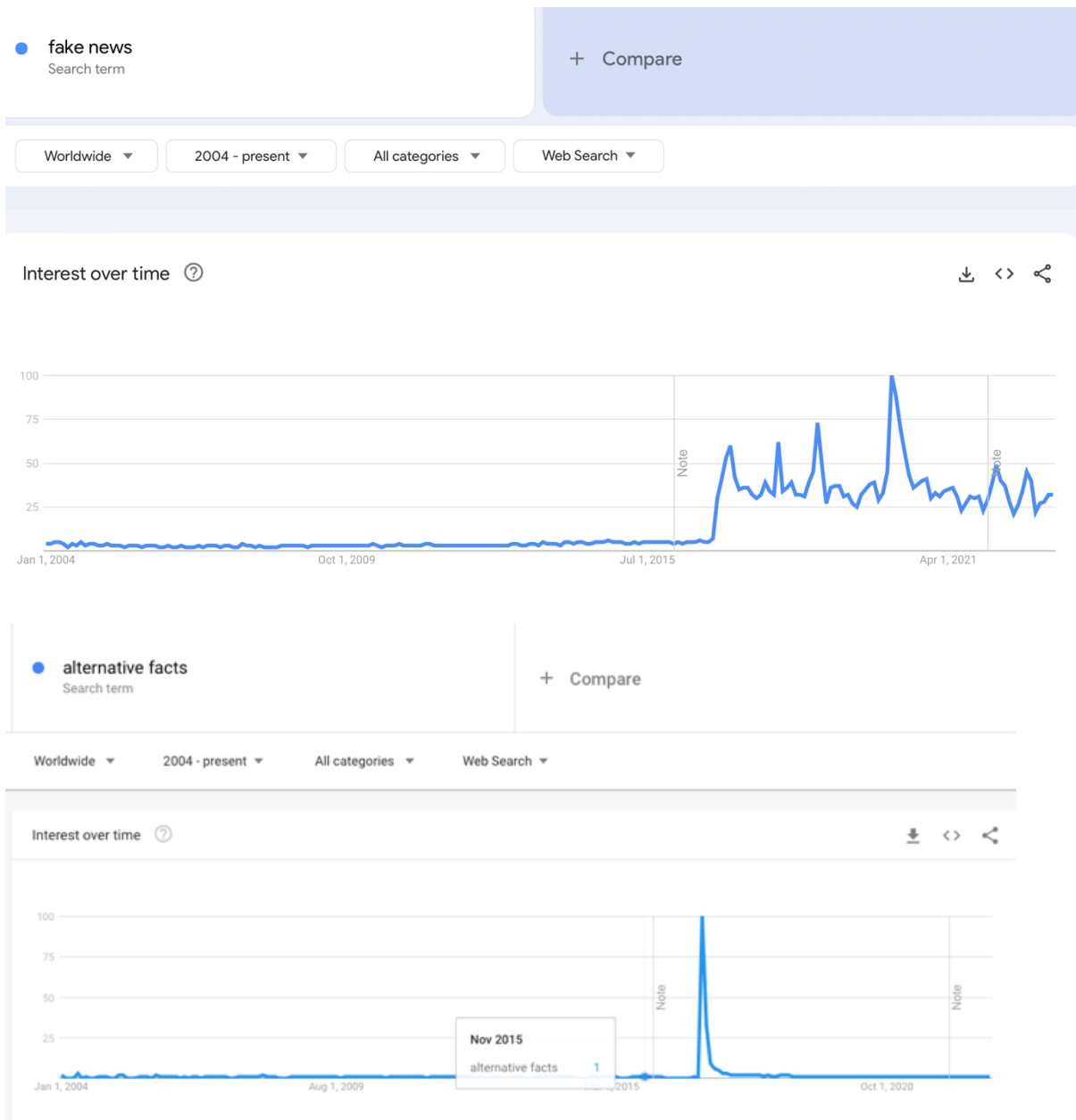
Figure 2 - Google search trends for popular terms related to disinformation (e.g. fake news, alternative facts") which spike in 2016, correlated with the 2016 United States Presidential election

Western nations have also historically used the spread of false information to maintain geopolitical and economic stability. For example during World War II, the British aired a radio show in Berlin pretending to be a young German who was revealing mismanagement by the Nazi government (e.g., an outbreak of diptheria) to weaken the regime's hold of Germany (Shaer, 2017) - even though the specific mismanagement never happened or not to the reported scale. In the 1950's, the CIA spread false information characterizing ousted

Guatemalan president Árbenz as a coward (though the CIA referred to him as "brilliant" when their goals were aligned with supporting Árbenz) in order to maintain the USA's economic stability and fight the feared global spread of communism (Ferreira, 2008).

However, like many aspects of an increasing digital society, the internet has now changed how disinformation can be created, spread and entrenched. While previous disinformation was spread through word of mouth, print or radio – and often through powerful individuals – digital tools allow disinformation to be created, tested and spread at much higher volumes and velocities, by whomever has the intent. In the 90s researchers were already concerned with how quickly images could be created and manipulated with bad intent and then shared and stored permanently without context for future permanent access on the internet (Floridi, 1996). Since then, disinformation has taken on many more forms from spam e-mails phishing for bank account information to the invention and use of memes to spread disinformation and ideas (Yankoski et al., 2021) – for example, a meme was released during the 2016 United States election to suggest that a policy in Hilary Clinton's platform was to specifically draft women into the military (Donovan, 2019).

In addition to new forms and increased volume of disinformation, the internet also provided for an unprecedented level co-ordination in disinformation creation and spread. Researchers have been able to track how individuals can - at an exceptional pace - create multiple social media accounts, create clout-building relationships between these accounts and entrench these accounts in active communities through targeted profile building. Then, using automation with some human intervention, these accounts can co-ordinate the timing of sharing disinformation among all these accounts to create waves of disinformation and subsequent waves of reinforcement to amplify these false messages (Ng & Taeihagh, 2021) to

create social disruption. By design, these increased capabilities have enabled an intensive fail-fast strategy that bad actors can use to generate highly effective disinformation, producing high-volumes of coordinated message to let the best-adapted form of a message win. A recent and notable example of this impact is the attempted insurrection of the United States government on 6 January 2021, driven by disinformation spread by "bot" accounts on Twitter.

Disinformation and its impacts on individuals and communities

Unfortunately, most individuals cannot tell the difference between facts and disinformation when they are presented in similar ways, which makes the increased volume, efficacy and spread of disinformation a significant challenge. For example, false information presented as news (e.g. with the aesthetic of a news site, self-labeled as news) triggers the same neural activity in individuals as when they read more traditional news publications, suggesting that it is the visual presentation of the news that creates trust versus the legitimacy of the source itself. In addition, in a high-information world, the least cognitively challenging ideas are taken up first, meaning that information competes and survives not based on its veracity, but on its simplicity and ability to be understood (Pech, 2003). Even more challenging, disinformation can modify human neurology and behaviour. For example, disinformation can manipulate memory recall, changing an individual's memories of past events to create a sound and logical path to justify behaviours and belief in current disinformation (Greenspan & Loftus, 2021). Individuals may also be unable to identify that their behaviours and attitudes have changed, reinforcing their current worldview as a long-standing truth, suggesting that the impacts of disinformation are also more systemic, subconscious and neurological in nature (Bastick, 2021). In one study, researchers were able to convince individuals that they had seen Bugs Bunny during a trip to Disneyland through

memory subversion via false advertisements the researchers created and shared with research participants. Even though the character and the tourist destination are owned by two completely different, separate and competing companies, researcher participants could not explain how their memory could be possible, but were insistent on its veracity (Loftus, 2005).

This manipulation of memory can create a particularly challenging causal loop (**Error! Reference source not found.**) for the pursuit of preserving of true and factual information and combating disinformation. As disinformation can manipulate memories, it can also shift the worldview of individuals, making them less resilient and more susceptible to similar disinformation. Conversely, these individuals build more resilience and resistance to any facts that contradict the narrative of disinformation, further supporting this reinforcing loop.

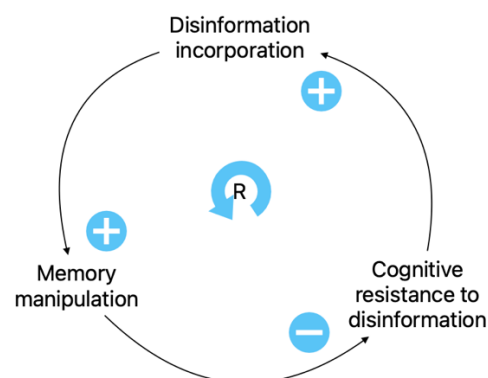*Figure 3 - Causal loop diagram demonstrating how consumption of disinformation reinforces openness and reduce friction to incorporating further disinformation.*

However, as humans are a communicative and social species, we have the ability and the drive to share information that we know and are convinced is interesting and true. This provides an opportunity for disinformation to spread and change collective memories and

belief systems in addition to individual ones. This can be witnessed through the development

of memes. While memes are more commonly understood to be a series of funny images, they

have demonstrated themselves as an effect method to spread disinformation (Donovan,

2019). The term meme was originally coined by Dawkins in the book, *The Selfish Gene*, as a

"Mimeme,…a noun that conveys the idea of a unit of cultural transmission, or a unit

of imitation" (Dawkins, 2006). The idea is that a meme is similar to a gene – a way

information can be expressed, shared and duplicate. However, instead of relying on sexual

reproduction and proteins, memes relied on information-sharing and culture to spread. This

viral nature of a meme – namely, its ability to spread quickly in communities and populations

– is also seen in social contagion theory. Social contagion theory hypothesizes that

impassioned speakers – those who speak with correctly calibrated emotion to improve

authentic connection - can embed ideas into and modify the memories in the listeners,

whether that idea is fact or fiction (Hirst, 2010). As this idea spreads and more impassioned

speaks are "infected" with these ideas, the further and faster the spread of the information

will grow.

Interestingly, this viral or memetic spread of information and ideas is not exclusive to

humans. For example, flocks of blackbirds will collectively identify a creature as a threat

over time as individuals within the flock watch each other's behaviours and reactions to a

new creature to arrive at a collective culture and understanding. Over time, both the blackbird

that encounters the new creature originally and the more passive observing blackbirds

demonstrate the same neural activity in response to the creature (Blackmore, 2006). However,

while this example can be seen as an adaptive behaviour to encourage collective survival and

fitness in blackbirds, the manipulation of this memetic transmission by disinformation has led

to more maladaptive behaviours in human populations and impacted our ability to manage

political polarization and maintain social cohesion. For example, while individuals are motivated to learn more about health when it is regarding their own survival, they unable to discern between good and fake news (Swire-Thompson & Lazer, 2019). As a result, individuals globally are building unique and individualized understandings of health and how it should be managed. This creates camps and groups of individuals who have deep-rooted by very contrasting opinions on health, making co-operation on global health events, like a pandemic, difficult.

In some cases, memes have been so successful in its ability to spread that it has created a new sort of language, creating a new form of digital division. This can be seen in the creation of the "metameme" (Figure 4) which frequent social media uses can use to communicate emotions and concepts, but excludes those who are less digitally-savvy. This challenges our ability to maintain consistent and effective world views and communications between audience segments – those that follow memes and those that do not.

Figure 4- Left: an original scene from the Lord of the Rings film has emotionally emphasizes how one cannot do a task as simply as they think. Centre: the conversion of that scene to a meme to emphasize how one cannot simply do a specific task. Right, a meme that mimics the first meme (centre) that makes a comment about lazy meme creation. However, the creation and understanding this second meme requires a cultural adoption and understanding of the original – this suggests that the first meme has hit "metameme" status.
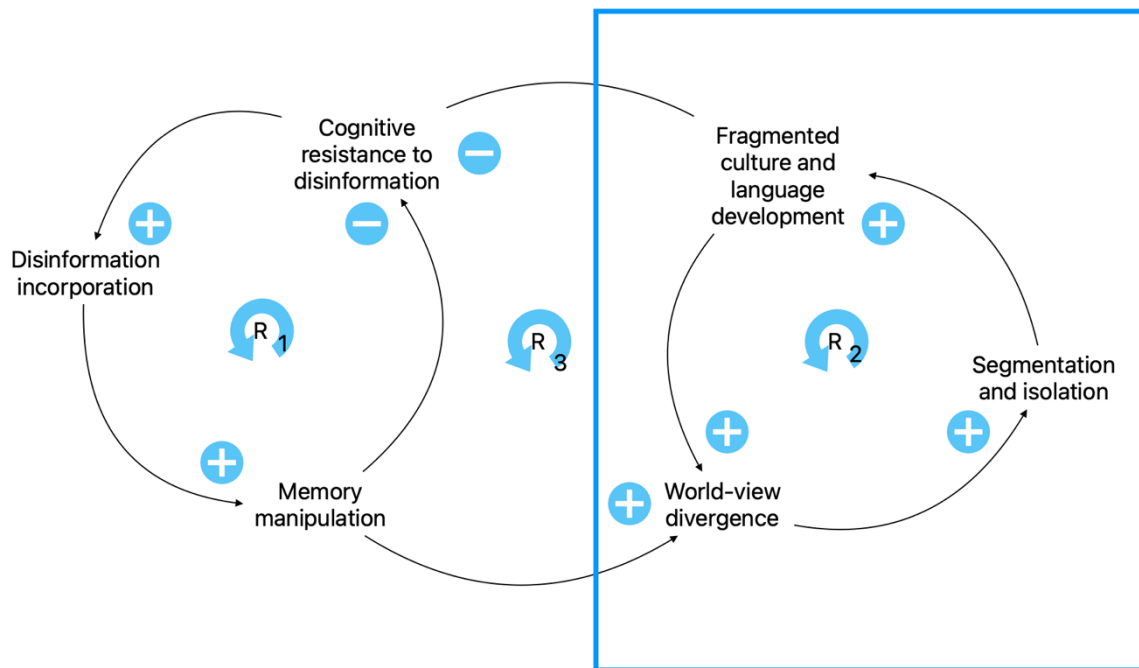
Figure 5 - Casual loop diagram demonstrating the community level impacts and effects of disinformation on cultural divergence and reinforcement.

These community-level impacts of disinformation and its ability to create separation in worldview and communications methods can further increase the feedback loops and impact of disinformation. This can be visualized in an expanded causal loop (Figure 5). In communities, the reinforcing loop of memory manipulation on the uptake of disinformation (R1) is further reinforced by cultural and language divergence in society (R3). Memory manipulation from disinformation can further drive divergence in world reviews and individuals in communities begin to recall past events differently, specifically in ways that reinforce a specific worldview (Liv & Greenbaum, 2020). This further segments and isolates individuals and communities, reducing the diversity of thought and perspectives (R3). These more individualized communities then develop a different, but more homogenous culture and language and communications styles that may be specific to them (R3). With less diversity of

thought and cognitive friction as a result, these communities become more susceptible to disinformation that nudges them to become more extreme in their worldviews (R3).

Regulation of disinformation

Given its harmful impacts of disinformation and its spread on a collective functioning society, there have been many attempts to regulate the generation and dissemination of disinformation. Regulations seem to be focused on two parts of the causal loop diagram (Figure 6): the first (F1) is focused on preventing the fragmentation of isolation of society by limiting the generation and dissemination of disinformation, while the second (F2) focuses on increasing cognitive resistance to the incorporation of disinformation.
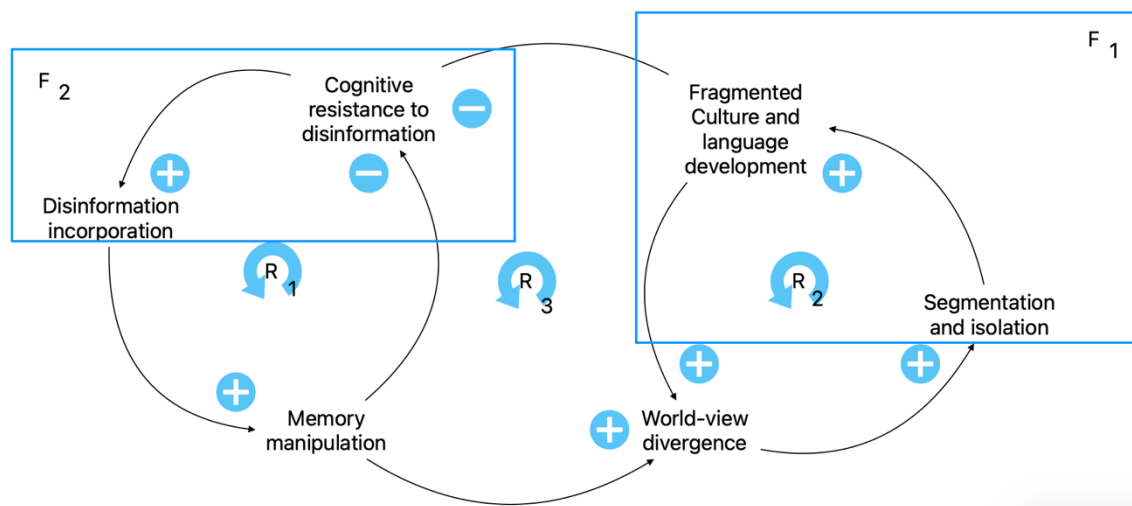
Figure 6 - This diagram highlights where in the causal loops current interventions against disinformation are focused. Focus areas are denoted as F1 and F2

On the first focus area (F1), institutions globally understand now that the volume and scale of disinformation being generated is an issue as seen through the pursuit of machine learning to regulate and control the creation and spread of disinformation (Choraś et al., 2021). However, regionally, there is higher variation in the philosophies and policies around the regulation of disinformation. For example, in the European Union, we see support for limitations on what can be expressed on the internet – from Germany's law which obliges

social media companies to remove unlawful content within 24hrs to the Czech Republic establishing an agency (the Center Against Terrorism and Hybrid Threats) which targets and counters fake news (Wood & Ravel, 2017). We see similar support for limiting regulations in Japan, Korea and Thailand though non-regulatory approaches (e.g. educational approaches) are preferred as a first line of defense in the first two countries (Cheng et al., 2021). Chile has implemented "brain rights" or "neurorights" in its constitution to protect its citizens from direct neurological manipulation (Guzman, 2022).

However, some countries are more cautious in their approach to regulation disinformation. In the United States, there is a high focus on how the regulation of disinformation challenges the First Amendment – more commonly referred to as the "right to freedom of speech". In essence, passing any laws that regulate disinformation requires proof that disinformation alone is responsible for friction in societies and that the definition of disinformation can be sufficiently narrow so as to not limit the creation and sharing for other information (Wood & Ravel, 2017). For example, in the United States Joe Biden had to pause the development of the Disinformation Governance Board – an effort to tackle disinformation on the internet – due to political interference and protests (Lorenz, 2022). Consultation with disinformation policy experts suggests that there is even less appetite, and therefore funding, for methods to build resilience against disinformation in individuals and communities. Similarly in Canada, regulation of disinformation – in particular during elections – took a largely deferential position, prioritizing freedom of expression and non-regulatory approaches, suggesting that a more severe approach could be taken should the need arise (Karanicolas, 2019). Consultation with Canadian disinformation policy experts suggests that Canada is similarly tepid towards direct interventions to increase resilience against disinformation.

Building resilience to disinformation

Given the constitutional, legal and operational challenges to regulating and limiting disinformation, there is also research that focuses on building resistance to the consumption and incorporation of disinformation in individuals (F2). As discussed, research in this space often refers to interventions as "inoculation" or "pre-bunking," referring to their preventative intent – with their conceptual roots dating as far back as the 1960s where Willaim McGuire observed how exposing individuals to disinformation and subsequently, counterarguments to the disinformation could increase individual resilience against disinformation (McGuire, 1961). In more recent studies on inoculation against disinformation, Sander Van Linden, Leiserowtiz and Maibach (2017) found that disinformation can nullify belief in previously stated facts, but announcements and warnings of disinformation help reduce this impact. Zhang (Zhang, 2021) also found that individuals who were exposed to clearly identified and labelled disinformation warnings were less impacted by disinformation and became more skeptical of information based on their sources.

Practical applications of resilience building can take multiple forms. For example in Estonia, after a significant cyber-attack from Russia in 2007, the government introduced a mandatory 35-hour disinformation course for those in the 10th-grade (Yee, 2022). Other researchers have started to create disinformation games as a form of resilience building and inoculation. In a game experiment by Roozenbeek and Sander Van der Linden (Roozenbeek & van der Linden, 2019), groups of students were asked to produce a news piece, assuming the role of a news producer with a desired angle. All groups are provided with a common set of facts and sources, but are asked to create a story using citations that aligned with the role's angle. Their preliminary findings suggest that such games help increase resilience to fake news in general,

but the nature of the experiment did not allow them to test against specific disinformation. In a similar vein, Dutch design studio – TILT Studios – has created two active inoculation games that ask the player to create and share disinformation with the goal of preventing a specific outcome (e.g. the construction of a Cat Park) or to sew general discord (Tilt Studios, 2022).

While regulatory interventions to disinformation are met with constitutional challenges or negative public opinion, there not as much insight, information or data on the public's receptivity to interventions that help build resilience to the consumption, digestion and impacts of disinformation. Participation in these interventions is often voluntary or executed at very small, local scales. However, as these interventions are often government-led or delivered via public institutions (e.g. schools), understanding public reception and acceptance of these interventions is important to scaling their delivery and their success. Given the challenges of regulating the creation and supply of disinformation, this project explores the reaction to future possibilities of interventions to increase mental resilience to disinformation and the public receptivity to these futures.

3. Methods

To assess public receptivity to future methods to build mental resilience against disinformation, this project had to define key drivers that should shape these futures and then enable research participants to experience them in an immersive way to enable effective data gathering. To accomplish this, the project was split into four parts:

  i.   Expert interviews and literature review on disinformation;

  ii. Driver and trends analysis;

  iii. Futures scenario creation and experiential futures; and

    iv.   Data analysis and comparison to literature

Expert interviews and literature review

       Interviews with experts and literature reviews were used in order to better understand the landscape, drivers and trends surrounding disinformation and resilience-building intervention. Literature reviews were completed to understand how disinformation is created and spread, and its impacts on individuals and communities. The review was also used to gain an understanding of the history of resilience-building interventions and experiments with current methods. Expert interviews with policy experts were used to understand how institutions currently think about building resilience against disinformation and the methods and interventions they have tried or are interested in. Experts were sourced from key authors in the literature review, expert recommendations and individuals steering the democracyXchange conference, where the experiential futures were installed.

Driver and trends analysis

From the literature and interviews, key trends were discovered regarding disinformation and the environment around its interventions. These trends covered the social, technological, political and values portions of the STEEP-V framework. Key trends that emerged from the literature review and expert interviews included:

STEEP+V Category	Trend
Social	**Digital nationalism** The generation of and reactions to disinformation reveal new sets of identities, communities, and social bonds outside of countries, ethnicities or current religions and secular institutions. For example, fans of the Korean pop music band "BTS" call themselves "ARMY" and will often join together to vote for the band in contests, influence business decisions and also swarm

	on disinformation about the band (Lee et al., 2022). Recent studies on media consumption and disinformation during the Ukrainian-Russian war discovered Europe and the UK were more likely to read and consume media from Russian sources, suggesting openness to ideas, disinformation and identity outside of their own national borders (Kling et al., 2022). This suggests that individuals, feeling part of a larger community, will participate in the generation, consumption and offensive against disinformation without direct government intervention.
Technology	**Digital arms race** Given the challenges of disinformation, governments are actively involving the use of AI and machine learning in policy discussions (Berkowitz, 2020) and decision-making (Meissner & Keding, 2021) to better understand its impacts. This also helps governments better understand and sensitize to how these technologies can be used to generate information and influence decision-making. For example, the Finnish government has established a Committee of the Future that helps the government identify future themes and explore policy routes and alternatives in preparation (Parliament of Finland, 2018). The United States Department of Defense also continues to operate the Defense Information School which trains military personnel in public affairs, social media, journalism and public broadcasting (Department of Defense, 2020). This represents an interest from governments to better understand problems like disinformation and become more active in a regulatory, educational or advisory role.
Economic	**The New PR** Experts reported that private companies are hiring content moderators from social media companies (e.g. Meta, Twitter) to help monitor and report false information about the company or help manage "bad news"

	and "damage control". Companies expect that these ex-social-media-employees will be able to seek out content that is bad for the company and work with tech companies to remove or moderate this content. This represents new investment from companies in trying to manage and maintain their brand with more direct invention in digital spaces instead of simply buying ads and running marketing campaigns. This form of brand-focused content regulation introduces another force outside of governments, social media companies, and individuals and communities.
Political	**Infiltration of the digital body** This trend represents the continued violation of the boundaries of the digital self. Individuals are slowly losing control over their digital profile: deepfakes can show an imitation of an individual doing and saying something crass and violent (Liv & Greenbaum, 2020). There is also more direct delivery of disinformation to users, from disinformation campaigns via voice messages in India and Lebanon (El-Masri et al., 2022) to "hacking" and manipulating virtual realities to force VR users to walk into walls (Casey et al., 2019). Anticipating the emerging boundaries of digital rights and privacy, Chile's government is the first jurisdiction to amend their constitution to include "neurorights" (Guzman, 2022) to protect mental privacy and manipulation by neurotechnology. **Rush to regulate** As disinformation continues to spread and create challenges, governments have moved to regulate the internet, content and technology companies. For example, Germany introduced penalties for tech companies that did not remove harmful content. The EU more broadly has introduced personal rights and privacy legislation (e.g. GDPR) and recently banned access to the Russian Times and Sputnik (Kling et al., 2022) media sites due the

	Ukrainian-Russian war. There have also been more radical forms of internet control such as the recent "shutdown" of the internet by the Iranian government to control protests (Alimardani, 2022) – a pre-requisite to this form of shutdown is to centralize network connectivity under government control.

Drivers and their influence

Shaping these trends are drivers that more broadly affect the regulation of digital spaces, use and understanding of technology and how information is created, shared and consumed. Figure 7 shows the key drivers that are underlying the trends identified in the above section and how they impact the methods and interventions that are created to build resilience against disinformation. These drivers are critical to review as they influence how disinformation is distributed and consumed and who may have control in building resilience against disinformation (Table 1).

Table 1 - Drivers and their areas of influence on the impacts of disinformation and the who influences how resilience against disinformation is built

Area of influence	Drivers
Disinformation creation and impact	Information decentralisation Privatisation of public space
Building resilience	Government regulation

For example, the rush to regulate, digital arms race and digital nationalism all are impacted by government regulation and their perceived dominion over the internet and information sharing. If these trends and drivers intensify, government may take a more direct, assertive and stronger role in building mental resilience in individuals and communities. In another example, increasing information decentralization drives a higher diversity of digital identities and communities, and the disinformation-sharing or disinformation-correcting efforts they choose to take. This could lead to more fragmentation in society's ability to pursue common and effective methods at building resilience against disinformation and make

more difficult the ability to keep-up with the novel volume and channels through which disinformation may spread. Lastly, the privatization of previously public or community-focused spaces may further drive individualization on the internet, enabling the increasing digitization of the mind and body via current and emerging technologies, making more of ourselves vulnerable to disinformation.
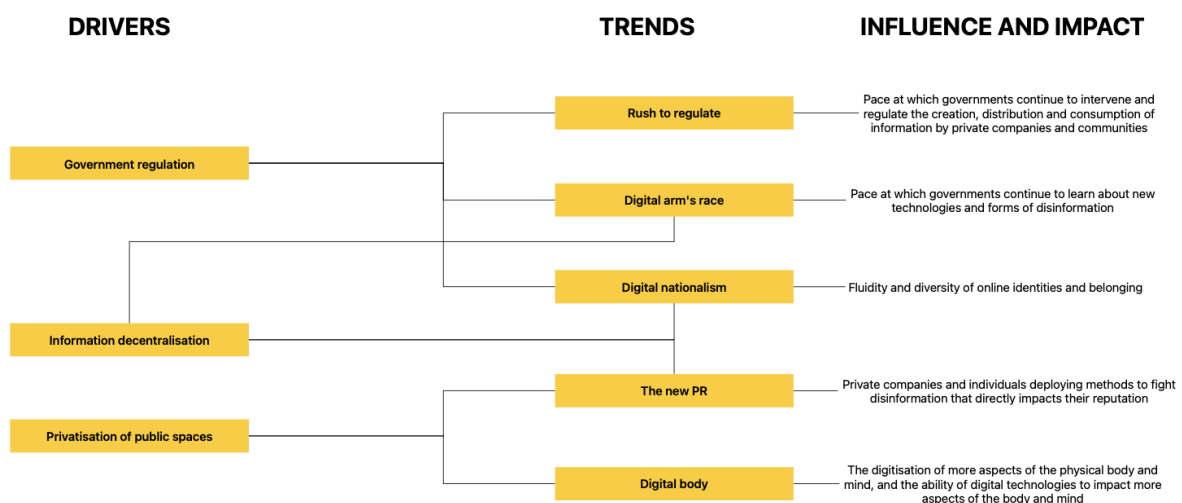
Figure 7 - Drivers behind current trends and their influence on disinformation creation, spread and mitigation. Drivers can influence multiple trends and trends can be influenced by multiple drivers.

Government content regulation

This driver represents the ongoing effort of governments to regulate access to the internet and content distributed on the internet. One polarity of government control includes disconnecting nations from the global network, eliminating internet connections and cellular signals or enforcing the use of a national intranet. On the other polarity, there is less government regulation of content and access; instead, the regulation of information is determined by the social media user base and culture of its users or its owner. Governments in these environments prefer to create educational initiatives in media literacy or rely on self-regulation by industry to government standards.

Information decentralization

Communications and content-sharing via the internet has fragmented the way information is generated, shared and agreed-upon as truth in a highly-digitalized world. Whereas information was previously declared as truth by newspapers, large media companies and academic journals, the emerging polarity follows a more networked approach. For example, users will conduct their own research or experiments and share the results via videos (e.g YouTube) or written content (e.g. Twitter, Reddit). Other users can comment directly or make content that analyzes the veracity or methods of the original post – readers or users can make their own decisions about the quality of the original study and information as a result. A similar process can happen where content creators take journal articles or government reports and deconstruct them with their personal expertise and opinions. In both scenarios, users respond to the clout of the content creator versus their associations with more traditional institutions – this creates phenomena like widely publicized conflicts between YouTube content creators.

Privatization of spaces

There has been increasing pressure to privatize the ownership over newly created spaces, from the physical to the digital. Since the 1950s, there have been discussions and debates about how previously public spaces in urban areas have become more privatized or commercialized, despite still being open to the general public (Devereux & Littlefield, 2017) More recently, interactions that would typically take place in public spaces now occur in virtual ones – from online gaming in MMOs to virtual meetings over video conference calls or chat platforms, these new digital spaces originated as private and commercial spaces. However, these spaces also provide more personalization as individuals can choose unique

usernames, avatars or special video effects – expressions that were not possible or less accessible in the physical and public realms.

Impacts and uncertainty of trends and drivers

Given their influence, the direction of these drivers creates a directionality for the emerging futures and may reinforce the mitigate the impact of other drivers. For example, the pro-activity of governments to learn more about new technologies and intervene with interventions to build resilience to disinformation directly influences the role of private sector players and the further privatization of public and community spaces. Increased information decentralization can further make it difficult to identify sources of truth, resulting in creasing identities to form and increasing the spread and use of disinformation. Further privatization of public and community spaces like the internet may drive increasing digitisation of the mind and body through the use of newer technologies (e.g. AR, VR and Neuralink) as demand for personalization increases.

To narrow the field of exploration for this study's experiential futures, the importance and uncertainty of these drivers of change were mapped to better understand their role in the futures of interventions in building mental resilience (Figure 8). Broadly, there was less certainty around how governments would regulate disinformation and develop policy or educational programs in future. As discussed, policy positions vary widely by country and political affiliation – ranging from stricter regulations and proactive interventions in Europe to a more laissez-faire and freedom-focused approach in North America. The government's ability to keep pace with new technologies and therefore, new ways of spreading disinformation is also uncertain, given the strong role disinformation has played in elections

over the past decade. However, the role of government continues to be very important given their role in regulating corporations, public institutions and society.

Boundaries and agency around the digital-self are also very important. As individuals continue to live in increasingly digital societies, more aspects of their body and mind will be digitized – for example, over the past two decades, people have moved from communicating on the internet via e-mail to providing tech companies access to their financial information, medical information and mental health data. However, there is high uncertainty about how individuals will be able to limit and control their exposure or resilience to disinformation as new digital channels continue to emerge before we can understand existing ones. For example, higher user awareness about privacy and personal data tracking by companies - and the consequent implementation of GPDR and a cookie-less future - are recent events in the internet's near three-decade life. Individuals and communities may not similarly struggle to understand and negotiate the risks of new technologies, as these tend to be emergent. In addition, these emerging channels provide new forums for individuals to form communities, and therefore create, share and spread information – this would increase the complexity of building mental resilience as interventions would need to accommodate ever changing methods of spread.

While the certainty of information decentralization and weaponization in building mental resilience interventions is high, its impact is less clear. Continuous growth in private and specialized technologies such as social media, wearables and digitization of daily services drives the further upload of personal data into different digital assets. This movement has enabled individuals to consume content and learn more about themselves through their platform of choice, even though companies are fighting for as much mindshare and data on

individuals as possible. With more decentralization, it is more difficult for societies to establish a common agreement on which sources of information are most reliable. However, this decentralisation would follow the trends and technologies established in the drivers and trends identified above – so while further decentralization definitely has a significant impact on the creation and spread of disinformation, its impact on building mental resilience is more tied to the technologies society accepts.
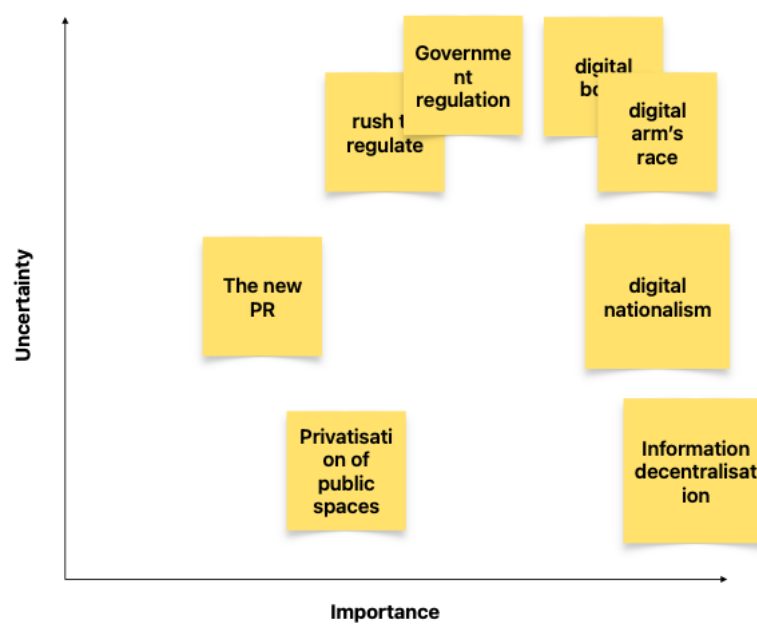
Figure 8 - Mapping of the trends and drivers around mental resilience development by uncertainty of their future and importance to implementation of mental resilience intervention policies.

Futures scenario development and experiential futures

Reflecting on the uncertainty and impact of these drivers, there are two drivers of change that emerge: government regulation and the boundaries of the digital-self. Combined, these forces focus on both the force and form of mental resilience interventions, the breadth of technologies they must consider, and their acceptance by communities. Plotting these drivers

of change on two axes (Figure 9), creates a matrix of four potential emerging futures around the development of interventions to increase mental resilience to disinformation for exploration. This 2x2 matrix method allows for the exploration of the intersection between the extreme polarities of these drivers to create distinctly different scenarios that can trigger a variety of participant experiences (Curry & Schultz, 2009).
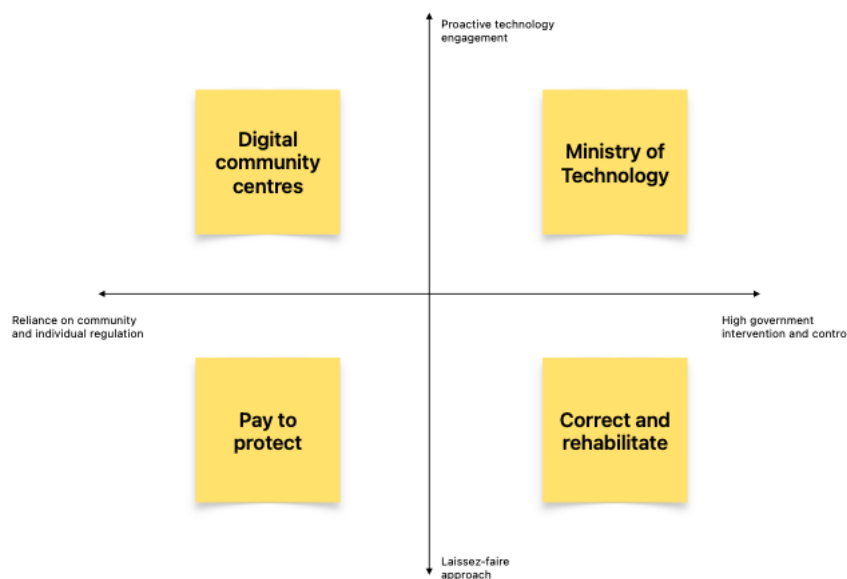
Figure 9 - A 2x2 matrix exploring the polarities of government intervention and proactive understanding of emerging technologies to create futures scenarios.

The axis on government regulation explores a polarity between more heavy-handed, direct interventions (e.g. aggressive content regulation, mandatory education programs, tight corporate regulations, etc. to a more laissez-faire educational approach, which relies on communities and individuals to participate in the development and roll-out of mental resilience interventions. Assertive government efforts can efficiently and quickly limit access to disinformation and develop and drive activities to build mental resilience, from sharing pre-bunking information to enforcing mandatory education programs. However, such strong

interventions are less effective or considered completely unacceptable in democratic countries. A more education-focused approach cannot limit the consumption of disinformation, but are more aligned and acceptable in democratic societies, driving uptake and maintaining the social stability required for these interventions to be sustainable.

The other axis focuses on the agency of individuals over their mind, body and data in digital spaces – from a more explicit ability to set boundaries in emerging and existing technologies to a less literate and cautious approach to embracing and using new tech. An example of an explicit boundary is the recent ability of users to discretely set specific data privacy restrictions and screen time limits on their mobile phones, increasing individual sovereignty over their personal data and behaviours. This is contrasting with the uptake of health technologies – from genetic screening services (e.g. Ancestry, 23andMe) to health monitoring services (e.g. continuous glucose monitoring) – driven primarily by curiosity and convenience with less consideration for the risks of digitizing this highly personal information. If society carries recent lessons learned forward, new technologies may reach mass adoption paired with improved regulations, education about how bad actors may leverage the technology to spread disinformation and how to identify and resist new forms of disinformation. In other futures, society may continue to lean into technology without hesitation, using critical events to reflect on the negative impacts of new media and channels.

Exploring the interactions between these two axes, four scenarios were developed for this study and participant testing: Ministry of Technology, Digital community, Pay to protect and Correct and rehabilitate. Of these four scenarios, three (Ministry of Technology, Digital community and Pay to protect) were turned into physical installations – experiential futures - with audio-visual elements at the democracyXchange (Figure 10). These three scenarios were

selected as they have the largest divergence from the probable future – Correct and rehabilitate – which is characterized by a strong dependency on government intervention even if the ability of this regulation to understand technology is significantly behind the speed of technological development.

The use of these installations and experiential futures is grounded in the philosophy of design fiction – a way to telling speculative stories through objects that help stretch the mind beyond the practicalities of science, but still grounded in the realities of science (Bleecker, 2022). This method enables participants to explore scenarios beyond the limits of what they see and hear day-to-day about the future (e.g. via the news, websites), through guiding them to stretch their imagination through the use of concrete artefacts created to specifically represent futures in more extreme and distant horizons. To effectively create experiential futures, the creation of these scenarios installations guided by the POEMS framework (Crawford, 2017), which involves carefully identifying how the objects, environments, messages and services used can reinforce the context and tone of each future.

Visitors to the democracyXchange were invited to visit the installations in between conference sessions. Installation visitors were asked to complete a short survey to assess their receptivity, preference and belief in each of the futures presented. 15 of the visitors that went through the installations provided feedback. The participants had various backgrounds, including work experience in government, design, non-profit, government relations and technology.
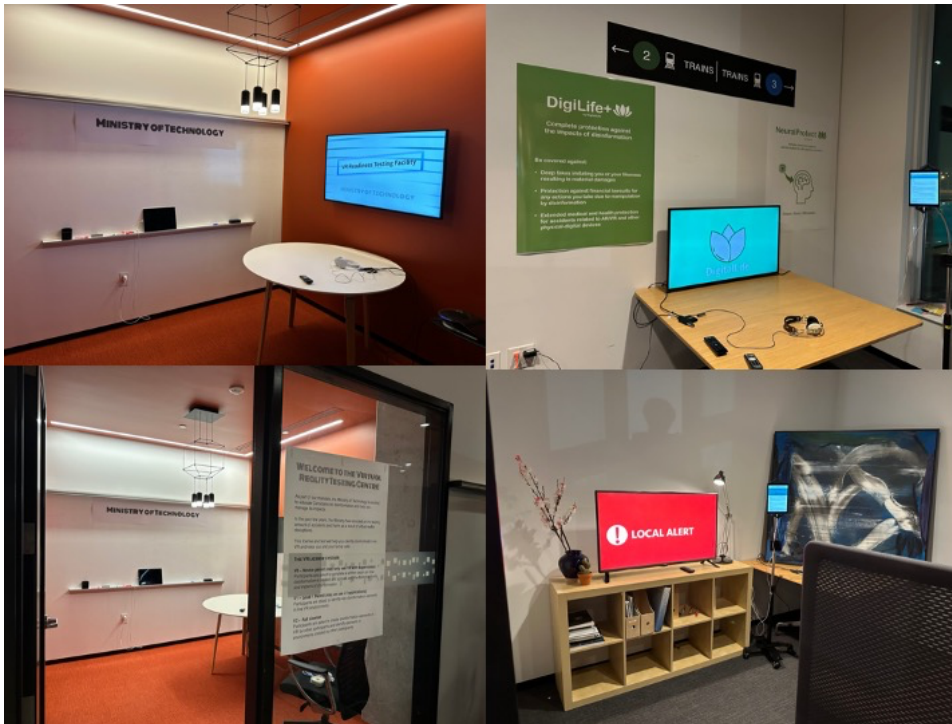
Figure 10 - Photos of the experiential futures installations at the democracyXchange. The scenarios represented are: Ministry of Technology (top and bottom left), Pay to protect (top right) and Digital community (bottom right)

1. Ministry of Technology

This future is characterized by high direct intervention from the government to build mental resilience in a world more pro-active in exploring the impacts and effects of new technologies. In this future, the ways disinformation can spread in new technologies and disinformation is researched and understood before mass adoption is allowed to occur. As a result, an individual's mental resilience to these new channels must be verified by the government before they are allowed to use this new technology. An example institution in this future is the Ministry of Technology, a government authority that develops tests and licensing systems for the use of new technologies. To demonstrate this scenario, the following artefacts were considered:

Artefact Category	Artefact
People	The participant is a candidate for a license to use new technology
Objects	Banner that thanks the candidate for their vigilance – signalling the community

	benefit of each individual taking responsibility for build resilience to disinformation
Environment	A government testing facility, like a driver's license testing centre
Messages	Key messages include: • New technology is useful, but can be risky • Bad actors can use new technologies to spread disinformation • Being able to identify disinformation is critical to community safety • Testing and licensing helps protect individuals and communities
Services	Testing for a license to use new technologies

2. Digital community

This future is characterized by educational and community-led approaches (e.g. NFP, CSO) to teaching and guiding communities through new and existing tech and general information literacy. While the government is concerned, they are less heavy-handed and prefer to fund research and interventions that are created and driven by communities and public institutions. Therefore, mental resilience in existing and new technologies is primarily done through education and advice, in public schools, community centres and across media outlets. Individuals and families are advised about disinformation events, but building mental resilience to these events is left up to individual proactivity. To demonstrate this scenario, the following artefacts were considered:

Artefact Category	Artefact
People	The participant is an individual watching community news in their own living room
Objects	A television/streaming news feed of disinformation events globally and locally
Environment	A living room environment, in the comfort of the participants home
Messages	Key messages include:

	• Disinformation alerts globally and how to manage daily tasks against these events • Local disinformation events and how to mitigate risks
Services	A news feed of alerts to warn individuals against disinformation

3. Pay to protect

This future is characterized by low government regulation and less pro-activity in trying to understand how disinformation is spread in new technologies. As a result, communities and individuals are trying to catch-up on how disinformation is spread in new technologies and protect themselves against mental resilience failures based on their personal interest and ability to pay. A key example in this future is the development of disinformation insurance – where individuals purchase insurance policies to protect themselves against the impacts of disinformation, but are offered discounts on their premiums if they elect to take educational classes or install neurological interventions so they themselves are more resistant to disinformation. To demonstrate this scenario, the following artefacts were considered:

Artefact Category	Artefact
People	The participant is on their daily trip and commute, seeing an ad for disinformation insurance
Objects	Disinformation insurance ads and advertisements for neural implants that prevent disinformation impacts
Environment	A subway/train station to signal the mass marketing of these interventions
Messages	Key messages include: • Disinformation is rampant and it is important to protect and insure individuals from its impacts • Proactive intervention (e.g. neural implants) is one method individuals can take to be safer and reduce the cost of protection

| Services | Disinformation insurance against deep-fakes or bad actions taken as a result of disinformation |
| | Neural implant to make individausl more resilient to disinformation |

4. Correct and rehabilitate

This future is characterized by direct intervention from the government characterized by regulation of older technology and dealing with the consequences of new technology – largely punitive in nature. As governments and communities are not pro-active in understand the impacts of technology and as a result, how to build mental resilience in an ever-changing digital world, their course of action is to persecute and a rehabilitate individuals and communities that cause material and significant harm to society from radicalization due to disinformation. In many ways, this is the future that is most probable reflecting on recent events like the attempted insurrections in the United States, Canada and Brazil, which were all driven by similar anti-vaccine, alt-right messages.

4. Analysis and Findings

Participant survey responses

Visitor responses indicated some clear preferences and aversions to different futures. The Pay to protect future, characterized by low government involvement and a lack of proactive technology engagement, triggered the most intense emotions (Figure 11) and was identified to be the least preferred future (Figure 12). Concerns around thought surveillance, physical implants and thought monitoring – even by private companies – were the primary concerns of this future. However, some participants did identify that this future felt realistic – primarily referencing the potential for profit in insurance as the reason for its realism. Interestingly, a

few participants also saw the future of a pro-active government – Ministry of Technology - as over-regulation and invasive, inciting similar reactions to the Pay to protect future. The community-led and proactive future – Digital community – also triggered emotional intensity from some participants. In particular, the volume of disinformation and alarming format triggered anxiety about how to process all these events.

However, the Digital community future was also considered to be both the most likely and preferred future (Figure 13, Figure 14). Participants noted that this future was the simplest to ladder towards – as it leverages existing infrastructure and aligns with how participants perceive community responses to disinformation - warning about disinformation, instead of actively trying to regulate and prevent it. The future of proactive government regulation – Ministry of Technology – was seen as the least likely, as participants were not convinced that government could keep recent with technology and regulation.
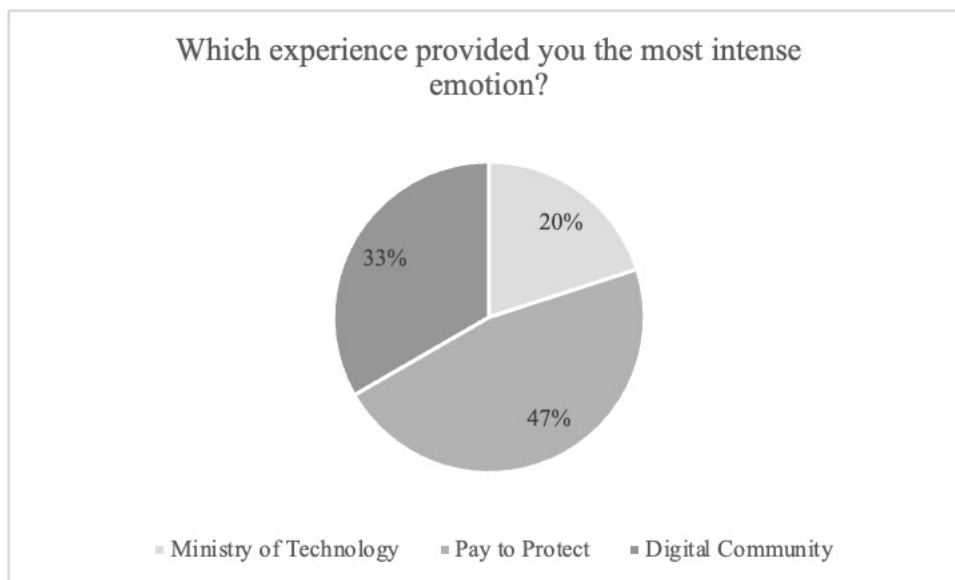

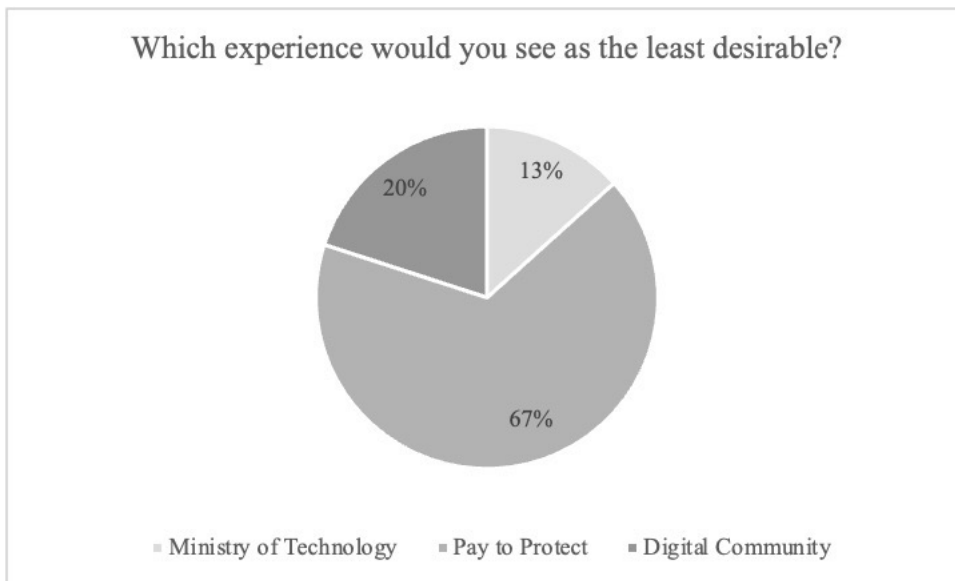*Figure 11 - Distribution on which scenario participants ranked as triggering the most emotional intensity*

Figure 12 - Distribution on which scenario participants ranked as being the least desirable
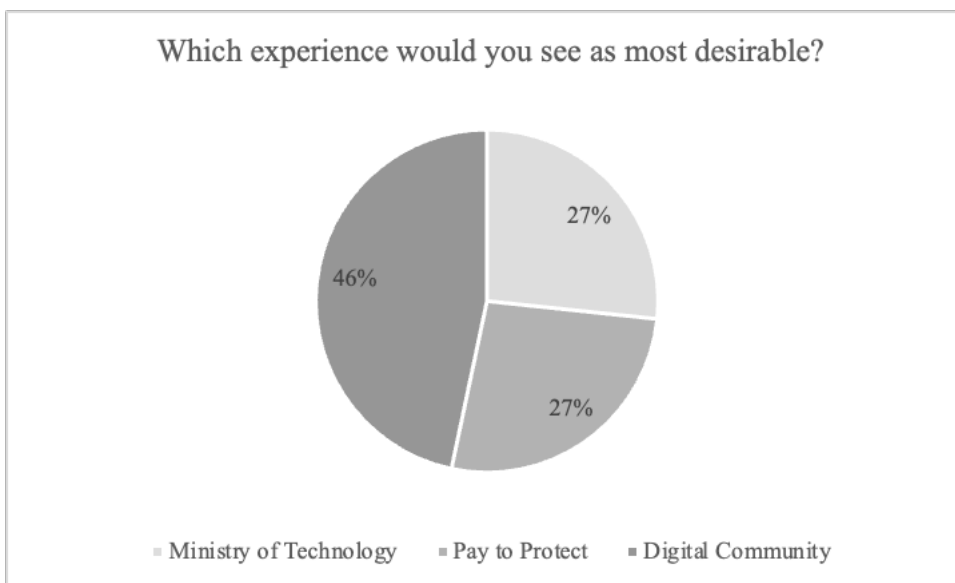

Figure 13 - Distribution on which scenario participants ranked as being the most desirable
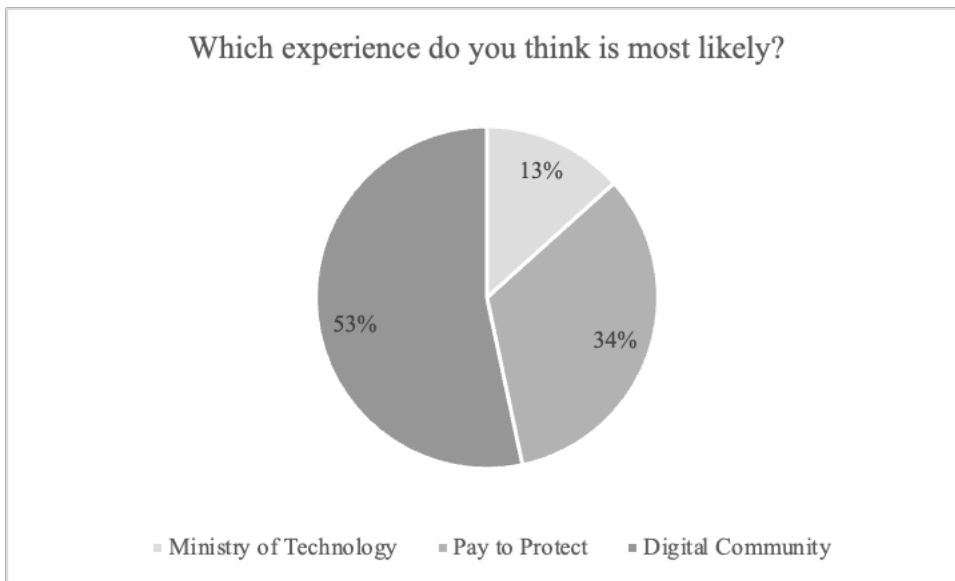
Figure 14 - Distribution on which scenario participants ranked as being the most likely

The tension between the government taking actions beneficial for communities-at-large and concerns over freedom of expression are consistent with the tensions over government regulation of disinformation generation – particularly in North America. As a result, government regulations to help individuals and communities build resilience against disinformation may be perceived to be both potentially behind the technology curve and highly invasive, impacting the legitimacy of any attempts. Therefore, governments that may want to be more direct in developing policy that helps build resilience against disinformation will need to demonstrate a strong grasp of the technological, psychological and legalistic qualities of the issue, demonstrating a high level of competency on the dynamics of how disinformation is shared and digested.

A more community-focused effort – like Digital Community – are unsurprisingly preferred as they represent the grey-space in-between government and industry control. However, similar to community preferences and divides over the veracity of different news and media sources, the legitimacy of who provides training and warnings against disinformation will need to be

carefully crafted. Even when immersed in a more proactive future, one participant responded to the Digital community scenario, "I thought the alerts were the fake news".

Findings and comparison to literature

Overall, the preferences and sensitivities of participants to the different futures are aligned with research and findings in the broader discourse on managing the spread and impacts of disinformation. Namely, individuals are concerned - even when it comes to building resilience against disinformation – with the amount of control government has over individual freedom of thought and expression. This is similar to the concerns individuals over government regulation of social media content to prevent the spread of disinformation as identified by Pielmeier (Pielemeier, 2020) and Tan and Sijie (Tan & Sijie, 2020). The participants' general preference for the Digital community scenario also supports a more grassroots effort and exploration of ways to build resilience against disinformation – like the more educational experiments and initiatives (e.g. disinformation games) explored by Roozenbeek and van der Linden (Roozenbeek & van der Linden, 2019). However, as these community programs scale, government endorsement or efforts that are too polished may draw individual suspicion and concern.

There are fewer direct parallels for how private industry may participate in the prevention or insurance against disinformation. However, looking at broader services around digital protection can expose some potential analogues. With an increasing number of accounts required to use online apps and services, password management has become an emerging issue. In the US, 41% of online adults reuse passwords (Olmstead & Smith, 2017) and in Canada, this number is as high as 55% (Anaya, 2021). As a result, security professionals have championed the use of password management tools as insurance policies against

security incidents and to protect private and sensitive information. Some Canadian telecommunications companies like Telus have begun to offer password management as part of their products and services as their consumer base becomes more digital (*Online Security Prevention | TELUS*, 2023). However, adoption of these tools has been varied – with non-users lacking literacy in why password management is important or – for those who are more digitally competent - a lack of perceived value of using the tool against the time and investment required to set up and maintain a password management tool (Fagan et al., 2017). Similarly, though VPN tools have become more prevalent in online marketing, their adoption remains restricted by price sensitivity and the perceived value of the protection offered (Ghobakhloo et al., 2011). As a result, though this future seems likely to participants – in that private industry will try to fill niches where they see profitable opportunities – public receptivity and adoption is not guaranteed.

Experiences using experiential futures

Using experiential futures through art installations enabled participants to better imagine and identify emotions identified with the broader environment. For example, in the Digital community future – some participants noted the safety and reassurance provided by the living room setting. However, this requires that the installation establishes context as clearly as possible, requiring in some cases very specific reference objectives that may not be as easily obtained. For example, the Pay to Protect scenario had a train banner to suggest that the videos and posters are advertisements. Some participants were confused by this asking "Why trains?" suggesting that the use of televisions for the videos and unframed posters, did not reinforce enough the idea of a public transit environment.

Driving engagement in the scenarios can also be difficult, installations with interactive elements can be intimidating or confusing to potential visitors. Convincing conference attendees to participate required direct asks, invitations, and attention. Pro-active expectation setting and signage can reduce social friction and manage expectations for participation. Installations are also resource-intensive process not only in preparation and set-up but also during visits and participation. Installation visitors often have questions about the purpose of the experience, and need very clear instructions and troubleshooting and guidance to complete the survey.

Areas for further research

This study was able to help signal where there are preferences and resistance to different and extreme scenarios of how societies and institutions may build resilience to disinformation in the future. However, they did not address the efficacy of the methods proposed in the scenarios and any changes to perception or acceptance of interventions that may result from seeing positive impacts. Knowing that more community-focused futures are preferred and that more community-focused and educational initiatives are demonstrating efficacy, it is important to explore perceptions around how these methods can scale. This includes the demonstrating the necessary resources required for their growth like government funding and endorsement or private industry resources, to gauge for potential friction points and acceptance. This can be achieved through using a narrower scope for experiential futures.

This study mainly used audio and visual elements to communicate the impact and environments of the future, primarily due to time and resource limitations. While these methods were able to help individuals identify the differences between the futures and provide different responses and reactions to these differences, there are many technologies

that can make the experiences more immersive. Repeating these installations and gauging participant responses with newer and more immersive technologies can reveal whether additional investment in technologies can create different, or more useful, results. For example, scenarios may be made more immersive using emerging mass market technologies like virtual reality, augmented reality and artefacts created from 3D printing.

5. Conclusions

There are many efforts focused on regulate the production and sharing of disinformation and these have been met with regulatory challenges and market realities (e.g. many disinformation channels are privately owned). Therefore, research has also focused on how individuals and communities can increase resilience to the impacts of disinformation. While there is early research on the efficacy of methods to build resilience, there is less insight on how these will be accepted by individuals or communities as they scale. This study set-out to use experiential futures to explore public receptivity to the different extremes of these possibilities.

From the data and analysis gathered via the experiential futures installations, public reactions to different extremes in building resilience against disinformation are similar to current tensions around regulating disinformation production and social media usage. Primarily, both government and private industry efforts are met with concerns over privacy, surveillance and control – however, participants observed the profit motivations driving private industry investments, even if they are not guaranteed to succeed. As a result, community-driven initiatives are generally considered the most positive and acceptable, but mostly due to their

perceived ease of implementation or their less invasive nature – there was little comment on their perceived efficacy.

Therefore, governments may not have an easier path to regulation through investments in building resilience, but they may be able to leverage their own institutions (e.g. education system) rather than convincing private corporations (e.g. social media companies). If governments pursue partnerships with community groups to build resilience against disinformation, a balance between efficacy and perceived control may become a continuous balancing act. Any initiatives that are more heavy-handed or direct to aim for efficacy may run into the same resistance and fear of control and surveillance. Exploring the futures in this more narrow scope of government-community may help reveal the right balance of government supporting, funding and community independence that is most acceptable in the scaling of effective solutions.

This project also explored the use of experiential futures in foresight work and gauge responses to different future scenarios. In this study, the experiential futures primarily used audio and visual elements in the form of posters and videos on televisions and screens. This enabled participants to identify the differences in the futures and as a result, their different reactions and reponses to each future. However, as technological advances help make new immersive and creative experiences more affordable and scalable (e.g. VR, AR, 3D printing), there is opportunity to understand whether investing time and resources in making experiential futures more immersive will generate different or more insightful responses and feedback from participants – and eventually – whether it provides reactions that are closer to reality.

References

1. Ahmed, W., Vidal-Alaball, J., Downing, J., & López Seguí, F. (2020). COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *Journal of Medical Internet Research*, *22*(5), e19458. Link to article

2. Alimardani, M. (2022, December 2). *OII | Grief, Technology and Revolution: A women, and youth led revolution in Iran*. Link to article

3. Anaya, T. (2021, June 30). *User authentication: 55% of Canadians reuse passwords across accounts*. GetApp. Link to article

4. Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, *116*. Link to article

5. Berkowitz, J. (2020, December 12). *The Evolving Role of Artificial Intelligence and Machine Learning in US Politics*. Linke to article

6. Blackmore, S. (2006). *Memes, minds and selves*.

7. Bleecker, J. (2022). Design fiction: A short essay on design, science, fact, and fiction. *Machine Learning and the City: Applications in Architecture and Urban Design*, 561–578.

8. Casey, P., Baggili, I. M., & Yarramreddy, A. (2019). Immersive Virtual Reality Attacks and the Human Joystick. *IEEE Transactions on Dependable and Secure Computing*, *18*, 550–562.

9. Cheng, J. W., Mitomo, H., Kamplean, A., & Seo, Y. (2021). Lesser evil? Public opinion on regulating fake news in Japan, South Korea, and Thailand – A three-country comparison. *Telecommunications Policy*, *45*(9), 102185. Link to article

10. Choraś, M., Demestichas, K., Gie\lczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. *Applied Soft Computing*, *101*, 107050.

11. Cook, J. (2019). Understanding and countering misinformation about climate change. In I. E. Chiluwa & S. A. Samoilenko (Eds.), *Handbook of Research on Deception, Fake News, and Misinformation Online* (1st ed., pp. 281–306). IGI Global. Link to article

12. Crawford, K. (2017, September 20). *Design Thinking Toolkit, Activity 3—POEMS*. Atomic Spin. Link to article

13. Curry, A., & Schultz, W. (2009). Roads less travelled: Different methods, different futures. *Journal of Futures Studies*, *13*(4), 35–60.

14. Dawkins, R. (2006). *The Selfish Gene* (9th ed.).

15. Department of Defense. (2020, January 21). *About*. Link to website

16. Devereux, M., & Littlefield, D. (2017). *A literature review on the privatisation of public space.*

17. Donovan, J. (2019, October 24). *How memes got weaponized: A short history | MIT Technology Review*. Link to article

18. El-Masri, A., Riedl, M. J., & Woolley, S. (2022). Audio misinformation on WhatsApp: A case study from Lebanon. *Harvard Kennedy School Misinformation Review*. Link to article

19. Fagan, M., Albayram, Y., Khan, M. M. H., & Buck, R. (2017). An investigation into users' considerations towards using password managers. *Human-Centric Computing and Information Sciences*, *7*(1), Article 1. Link to article

20. Fallis, D. (2014). The Varieties of Disinformation. In L. Floridi & P. Illari (Eds.), *The Philosophy of Information Quality* (pp. 135–161). Springer International Publishing. Link to article

21. Ferreira, R. G. (2008). The CIA and Jacobo Arbenz: History of a disinformation campaign. *Journal of Third World Studies*, *25*(2), 59–81.

22. Floridi, L. (1996). Brave.Net.World: The Internet as a disinformation superhighway? *The Electronic Library*, *14*(6), 509–514. Link to article

23. French, A. (2019). The Mandela Effect and New Memory. *Correspondences*, *6*(2).

24. Ghobakhloo, M., Arias-Aranda, D., & Benitez-Amado, J. (2011). Adoption of e-commerce applications in SMEs. *Industrial Management & Data Systems*.

25. Greenspan, R. L., & Loftus, E. F. (2021). Pandemics and infodemics: Research on the effects of misinformation on memory. *Human Behavior and Emerging Technologies*, *3*(1), 8–12. Link to article

26. Guzman, L. H. (2022, January 18). *Chile: Pioneering the protection of neurorights* [The UNESCO Courier].

27. Havelin, M. (2021). *Misinformation & Disinformation in Canadian Society A system analysis & futures study* [PhD Thesis].

28. Hirst, W. (2010). The contribution of malleability to collective memory. *The Cognitive Neuroscience of Mind*, 139.

29. Karanicolas, M. (2019). Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling'Fake News'. *Available at SSRN 3423092*.

30. Kling, J., Toepfl, F., Thurman, N., & Fletcher, R. (2022). Mapping the website and mobile app audiences of Russia's foreign communication outlets, RT and Sputnik, across 21 countries. *Harvard Kennedy School Misinformation Review*. Link to particle

31. Lee, J. H., Santero, N., Bhattacharya, A., May, E., & Spiro. (2022, September 28). *Community-based strategies for combating misinformation: Learning from a popular culture fandom | HKS Misinformation Review*. Link to article

32. Liv, N., & Greenbaum, D. (2020). Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. *AJOB Neuroscience*, *11*(2), 96–104. Link to article

33. Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366.

34. Lorenz, T. (2022, May 18). *Disinformation Governance Board 'paused' after just 3 weeks—The Washington Post*. Link to article

35. McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *The Journal of Abnormal and Social Psychology*, *63*(2), 326.

36. Meissner, P., & Keding, C. (2021, October 12). *The Human Factor in AI-Based Decision-Making*. Link to article

37. Ng, L. H. X., & Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy & Internet*, *13*(4), 560–585. Link to article

38. Olmstead, K., & Smith, A. (2017, March 22). *What Americans Knows About Cybersecurity | Pew Research Center*. Link to article

39. *Online Security Prevention | TELUS*. (2023, March 20). Link to website

40. Parliament of Finland. (2018, January 26). *Committee for the Future*. Link to website

41. Pech, R. J. (2003). Memes and cognitive hardwiring: Why are some memes more successful than others? *European Journal of Innovation Management*, *6*(3), 173–181. Link to article

42. Pickard, V., Zimdars, M., & McLeod, K. (2020). Confronting the misinformation society: Facebook's "fake news" is a symptom of unaccountable monopoly power. *Fake News: Understanding Media and Misinformation in the Digital Age*, 123.

43. Pielemeier, J. (2020). Disentangling Disinformation: What Makes Regulating Disinformation So Difficult? *Utah Law Review*, *2020*, 16799.

44. Prasad, D., & Bainbridge, W. A. (2021). The Visual Mandela Effect as evidence for shared and specific false memories across people. *Psychological Science*, 09567976221108944.

45. Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570–580. Link to article

46. Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), eabo6254. Link to article

47. Shaer, M. (2017, April). *Fighting the Nazis With Fake News | History| Smithsonian Magazine*. Link to website

48. Southwell, B. G., Thorson, E. A., & Sheble, L. (2018). *Misinformation and Mass Audiences*. University of Texas Press. Link to book

49. Swire-Thompson, B., & Lazer, D. (2019). Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health*, *41*, 433–451.

50. Tan, D., & Sijie, J. (2020). Fake News, Free Speech and Finding Constitutional Congruence. *Singapore Academy of Law Journal*, *32*.

51. Taylor, A. (2016, November 26). *Before 'fake news,' there was Soviet 'disinformation'—The Washington Post*. Link to article

52. Tilt Studios. (2022, March 24). *Serious Games*. Link to website

53. van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges (Hoboken, NJ)*, *1*(2), 1600008. Link to article

54. Wardle, C. (2022, September). *Understanding Information disorder*. First Draft. https://firstdraftnews.org:443/long-form-article/understanding-information-disorder/

55. Wood, A. K., & Ravel, A. M. (2017). Fool me once: Regulating fake news and other online advertising. *S. Cal. L. Rev.*, *91*, 1223.

56. Yankoski, M., Scheirer, W., & Weninger, T. (2021). Meme warfare: AI countermeasures to disinformation should focus on popular, not perfect, fakes. *Bulletin of the Atomic Scientists*, *77*(3), 119–123.

57. Yee, A. (2022, January 30). *The country inoculating against disinformation—BBC Future*. Link to website

58. Zhang, N. (2021). *Inoculating the Public Against Misinformation: Testing the Effectiveness of" Pre-Bunking" Techniques in the Context of Mental Illness and Violence* [PhD Thesis]. University of South Carolina.