UNIVERSITY OF CALIFORNIA, IRVINE

DoomScroll: Modding Among Us to Combat Misinformation

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Informatics

By

Garrison Akira Wells

Dissertation Committee:
Professor Constance Steinkuehler, Chair
Professor Katie Salen-Tekinbas
Professor Mia Consalvo
Professor Kurt Squire

2024

DEDICATION

To my parents, my siblings, Nat, Micah, and Randy,

Thank you all for trusting the plan.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Thank you to Constance Steinkuehler for her support, encouragement, and guidance throughout the past five years. You took a chance on me and I am forever grateful.

My committee members Katie Salen, Mia Consalvo, and Kurt Squire, as well as Stephen Schueller, who all gave me the trust and guidance to dream a little too big.

The Department of Informatics for generously supporting this work. To Ashley Chia Sun, Brandon Retana, and Emily Chavez for their contributions to data collection. And thank you also to Mark Deppe and Brenden Alvarez at UCI Esports for lending their time and resources to the completion of this project.

Agnes Romhanyi and Alaina Klaes, my partners on this project and without whom none of this would be possible. Alaina, your hardwork and enthusiasm was a constant reminder of what a joy it truly is to be studying games. Agi, I couldn't have asked for a better collaborator and friend to journey through this program with.

My friends and colleagues in the Games+Learning+Society Center, who listened to my rants and tolerated my tiara. Working with you all has been an absolute joy.

Randy's Boys, The Moon Rats, The Chemistry Department, OWO Watch, informatics folks, Kyle, Emory, Tyler, and all of the other friends I've made along the way, thanks for supporting your boy.

Natalie Minnetian, for giving me a home when I had lost all hope to build one.

And finally, my family, for their unending love. My parents for allowing me to pursue my passions, even if I haven't thought everything out yet. My siblings for always listening to my dumb jokes and watching my great dances. The dogs and cats for being sweet angel babies, too pure and good for this world.

VITA

Garrison Akira Wells

EDUCATION

2024	**Doctor of Philosophy in Informatics,** University of California, Irvine
2024	**Master of Science in Informatics,** University of California, Irvine
2017	**Master of Science in Psychological Research,** University of Edinburgh
2016	**Bachelor of Arts in Psychology,** University of California, Santa Barbara

PUBLICATIONS

Steinkuehler, C., Anderson, C.G., Reitman, J.G., Lee, J.S., Wu, M., Wells, G. & Gardner, R.T. (2023). *Enriched Esports: The Design and Four-Year Examination of a School-Affiliated Competitive Videogame Program for Youth*. Journal of Interactive Learning Research, 34(1), 59-119. Waynesville, NC: Association for the Advancement of Computing in Education (AACE). Retrieved June 5, 2023 from https://www.learntechlib.org/primary/p/221828/.

Squire, K., Wells, G., Anderson-Coto, M. J., & Steinkuehler, C. (2023). *Casual Games, Cognition, and Play Across the Lifespan: A Critical Synthesis.* ACM Games: Research and Practice. https://doi.org/10.1145/3594534

Wells, G., Romhanyi, A., Reitman, J. G., Gardner, R., Squire, K., & Steinkuehler, C. (2023). *Right-Wing Extremism in Mainstream Games: A Review of the Literature*. Games and Culture, 0(0). https://doi.org/10.1177/15554120231167214

Nour, N., Wells, G., & Steinkuehler, C. (2023). *Hireability Gaps and Frictions Between University Game Programs and Industry Hires*. International Game Developers Association.

CONFERENCE PRESENTATIONS

Steinkuehler, C., Wells, G., Romhanyi, A. UCI Student Exposure to Toxicity & Extremism in Online Games: The Problem, Its Consequences, and Possible Solutions. ASUCI Reclaim Mental Health Conference 2023. Irvine, CA

Wells, G. Susceptibility to Misinformation & Games as Interventions. Games Learning Society Conference 2022. Irvine, CA

Wells, G., Anderson, C.G., Steinkuehler, C. Three Years of NASEF Student Outcomes. UCI Esports Conference 2020. Irvine, CA

Yao, D., Reitman, J.G., Lee, J., Wells, G., Steinkuehler, C. CapturEsports: Data Collection for Teams Research. UCI Esports Conference 2020. Irvine, CA

ABSTRACT OF THE DISSERTATION

DoomScroll: Modding Among Us to Combat Misinformation

By

Garrison Akira Wells

Doctor of Philosophy in Informatics

University of California, Irvine, 2024

Professor Constance Steinkuehler, Chair

Misinformation is one society's most pressing issues, spreading division and chaos across the globe. Videogames have become one of the more promising mediums for misinformation interventions, teaching players common falsehood indicators and helping them develop their discernment abilities. This thesis details the design, playtesting, and evaluation of a digital game intervention intended to improve players' misinformation discernment abilities. Through a design narrative, it first describes the design goals and development of an initial iteration of *DoomScroll*, a modified version of the popular social deception game *Among Us.* It then reports on the playtesting of DoomScroll and how player feedback was used to refine and improve the mod's gameplay. It concludes by detailing a randomized control trial conducted to more rigorously test DoomScroll's efficacy as a misinformation intervention tool when compared to a control condition. Player feedback regarding its socialization and replayability suggests that DoomScroll did meet its design goals by improving player motivation and increasing the likelihood of repeated use. However, further analyses showed that the mod did not produce any statistically significant improvement in players' discernment abilities. The qualitative findings were more positive, suggesting an uneven distribution of process-driven learning did occur. This study contributes insights to improve misinformation game design and into the potential of modding as a medium for educational game development.

Chapter 1: Introduction

Misinformation has quickly become the most prevalent issue of our time, with the internet's now broad accessibility and lack of editorial oversight allowing bad faith actors to produce and share falsehoods that disrupt the social discourse and lead to a harrowing disconnect in what was once a collective shared reality. We have recently witnessed some of the repercussions of this content within the United States, as targeted misinformation campaigns influenced significant portions of the population to violently question the legitimacy of a democratic presidential election (Silverman et al, 2022) and express apprehension towards receiving scientifically-validated vaccinations, occasionally at the cost of their own lives (Brennen et al, 2020). Research has shown that the rise of misinformation increased significantly during the time period before and after the 2016 United States presidential election (Allcott et al, 2019). However, the reality is that these problems had been brewing under the surface for decades. From the "vaccines cause autism" conspiracy theories stemming from a single retracted study (Wakefield et al, 1998) to the laundry list of conspiracies that surrounded Covid-19 treatments (Enders et al, 2020), scientists have struggled to keep their findings from being misrepresented and their recommendations from being ignored (Jolley & Douglas, 2014; Iyengard & Massey, 2019; Loomba et al, 2021; Pluviano et al, 2017). And while the social media campaigns targeting the 2020 presidential election results were widely publicized and scrutinized (Sanderson et al, 2021), prior politically-driven misinformation efforts like the "Birtherism" accusations against Barack Obama (Berinski, 2012) and the unfounded threats of weapons of mass destruction (WMDs) in Iraq (Arsenault & Castells, 2006) show that it simply joins a long line of misinformation bred for political gain. The situation has become so dire to this point that the World Economic Forum (2024) has recently named misinformation as the greatest societal threat over the next two years.

Numerous interventions have been put forth to combat these efforts, including stricter user-generated content regulations on social media platforms (Einwiller & Kim, 2020), debunking mechanisms built into the platforms as well as third party fact-checking organizations (Pavleska et al, 2018), and media literacy campaigns intended to give individuals the tools to identify and avoid misinformation on their own (Dame Adjin-Tettey, 2022). While research has shown that all are effective, each method has limitations that prevent it from being truly impactful.

Recently, one potential intervention tool that has shown notable promise is games. Games have long been proven to improve learning and engagement (Garris et al, 2002; Squire, 2003; Steinkuehler, 2004; Steinkuehler et al, 2023) and have been deployed as effective interventions in areas like health (Lu et al, 2013) and wellbeing (Pallavicini et al, 2018). Through the process of inoculation, presenting players with a piece of information and tasking them with determining whether that information is true or false helps them adjust to potential misinformation techniques before coming in contact with them in real life. What's more, games allow for controlled parameters that other intervention methods cannot match, giving game/intervention designers the freedom to place players in environments that will more easily translate to situations they might face in the real world. However, games such as these are still in their infancy and few have had their impacts rigorously tested. Most contain single-player narratives that teach the player common misinformation techniques from either the perspective of the creator or the audience (Dejong, 2023), but lack the social elements that make misinformation discernment truly difficult (Basol et al, 2021; Roozenbeek & Van der Linden, 2020) or the replayability that can help boost up their defenses whenever the increased discernment begins to wane (Maertens et al, 2020). Many of these games have been tested for general player experience and self-reported confidence in their discernment abilities, but very few have undergone the strict protocols of a randomized control trial.

Research Goal

The goal of this dissertation is to develop and evaluate a new game-based misinformation intervention that addresses these issues by

1. Allowing players to continuously practice and develop their skills under novel situations,
2. Replicating the social context and pressures in which misinformation is usually faced.

In this work, I conceptualize the game's initial structure and learning mechanisms based on insights taken from the existing academic literature on misinformation and educational game design fundamentals. My development team and I then iterate on that design, adjusting different aspects based on player feedback in order to accentuate the educational features while improving the overall player experience. I then conduct a randomized control trial to rigorously assess the final build and determine whether it serves as an effective intervention tool. The results of this project have implications for the development of misinformation game interventions and how game design fundamentals might be better integrated into the process.

Dissertation Overview

Chapter 1 has discussed the context in which this research has been conducted and its overarching goals. Chapter 2 reviews the broader academic literature around this topic, detailing the characteristics of misinformation and its spread, the current theories of intervention, and how games have become a promising medium for future intervention efforts. Chapter 3 will outline the initial conception, educational objectives, and development of our digital game intervention. Chapter 4 describes our playtesting process and how player feedback influenced further iterations on our design. Chapter 5 reports on the randomized control trial conducted to assess the intervention's efficacy in improving players'

discernment when compared to a control group. And finally, chapter 6 provides a final summary and conclusions.

Chapter 2: Background & Related Work

Definitions

To begin, it is important to define the different types of misleading information and how they differentiate from one another. Baines and Elliott (2020) offer the following definitions of *misinformation, disinformation,* and *malinformation*, the three prevailing categories of misleading information, along with explanations standardized to their proposed conceptual framework for greater clarity. I also include Wu et al.'s (2019) definition for *fake news*; while fake news is considered a subset within misinformation rather than its own category (explaining why Baines and Elliott do not define it), I feel its popular usage in modern discourse necessitates defining it on its own here.

1. *Misinformation* is defined as "unintentionally false information". The message has no intention to deceive, but is not equivalent to the truth.
2. *Disinformation* refers to "deliberately misleading information". The message is intended to deceive and is not equivalent to the truth.
3. *Malinformation* is "reconfigured truth information". The message is intended to deceive, but is equivalent to the truth.
4. *Fake News* is "intentionally-spread misinformation that is in the format of news. Recent incidents reveal that fake news can be used as propaganda and get viral through news media and social media" (Wu et al, 2019, p. 81).

Under these definitions, *disinformation, malinformation*, and *fake news* all require knowing or assuming an author's intent when they produce misleading information (Baines and Elliott, 2020), while *fake news* also only covers misleading information produced within one specific structure. The content of

this dissertation will focus on general responses to false information rather than on any specific examples of the information itself, removing the need to speculate on any author's intent. Therefore, in order to better maintain its conceptual clarity and stay consistent with the term most commonly used in the existing literature, *misinformation* will act as the default term used going forward.

Despite the standing definitions above, the often fluid nature of the "truth" can make it difficult to formulate a concrete understanding of what constitutes misinformation. As Vraga and Bode (2020) note, "we often assume that misinformation is related to a fundamental underlying truth – something is either true or not – but to some extent, misinformation as a measured concept is dependent on the state of evidence, expert beliefs, and the information environment in which they occur" (p. 141). They point to how we evaluate both the experts and the evidence involved on a topic as major factors in determining whether the information being presented is true or false (fig. 1).

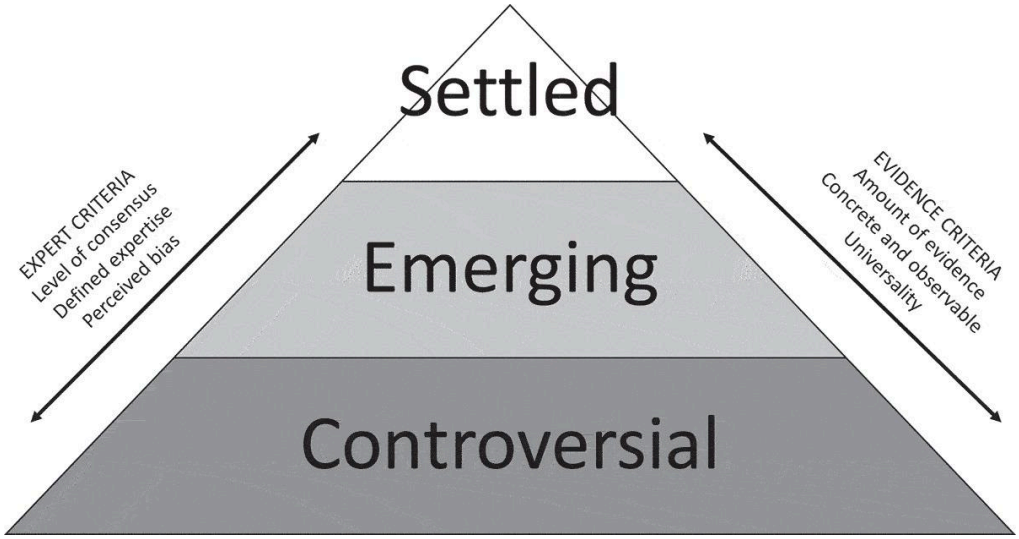

Figure 1. How we define factual information vs. misinformation is largely shaped by how we assess the experts and evidence available (Vraga & Bode, 2020).

Consensus among experts often appears as an objective standard for figuring out the truth; however, determining what qualifies as expertise and what level of agreement constitutes consensus is far more subjective. Experts from unrelated fields weighing in on topics beyond their purview (Readfearn, 2016) or with personal motivations to break from their peers (Brumfiel, 2021) can increase doubt and confusion. Moreover, even if the current evidence points to what appears to be an obvious truth, both the quantity and quality of the evidence available often changes over time (Fazio, 2021). To use Covid-19 as an example, early judgments were made regarding the efficacy of masks, and public health guidelines were set accordingly. Those guidelines have been adjusted as the state of the situation has changed and as more research has been published on the subject. While this ultimately follows the scientific process, speculative evidence can appear contradictory and sew distrust in expert opinions (Chang, 2015).

This is not to say that researchers must set strict parameters around what qualifies as "misinformation" in order to standardize its meaning. Rather, as Vraga and Bode (2020) suggest at the close of their article, they must simply be "more transparent in how settled the issues used are, and what criteria for selection are being used" (p. 141). Misinformation is far more complex than simple judgments of true and false, and so accounting for these complexities is vital to truly understanding it.

Susceptibility to Misinformation

Many different factors can impact an individual's level of susceptibility to believing misinformation (fig. 2). These factors influence everyone differently, to varying degrees, and often in tandem with one another. For instance, whether an individual believes or doubts public health guidance from the federal government can be based on their own understanding of the subject matter, their evaluation of a federal official's perceived expertise and biases, the individual's own political affiliations, or even just how often they hear the guidance repeated. With this in mind, I will organize this section by first describing the characteristics of misinformation that cause it to spread and multiply among

populations so easily. I will then go into detail about the cognitive processes and heuristics that, in the mind's efforts to efficiently sort out information, leave it vulnerable to these falsehoods. Finally, I will examine several societal trends and dynamics that exemplify how misinformation can be abused in practice.

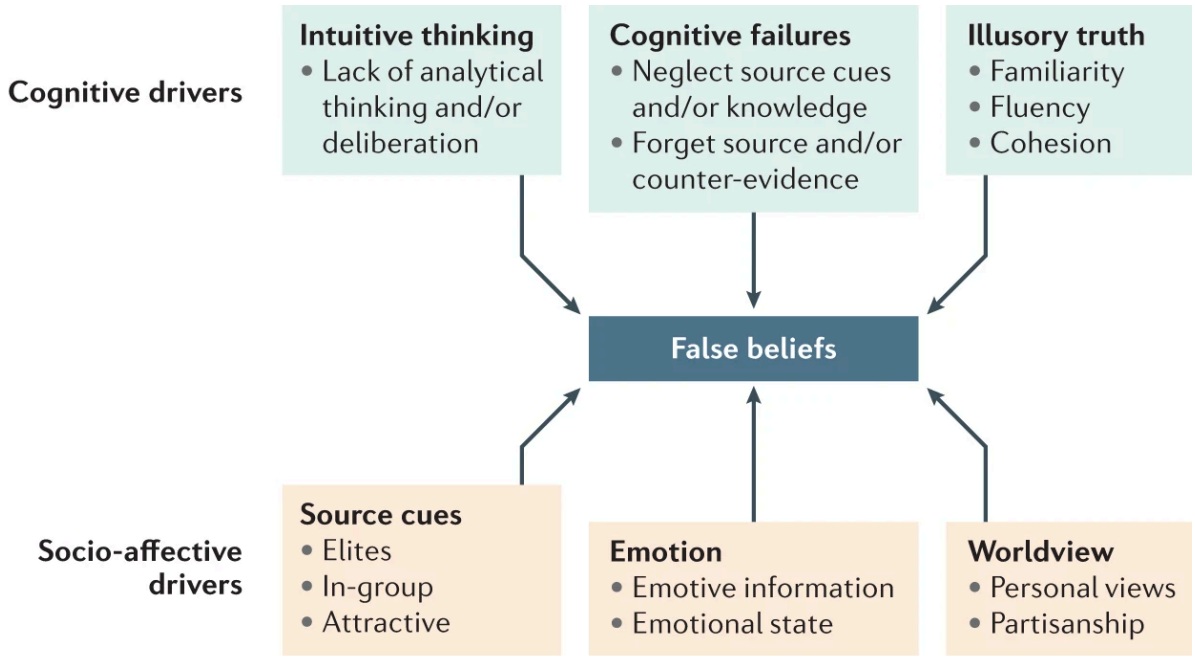

Figure 2. Broad categorizations of some of the drivers of false beliefs (Ecker et al, 2022).

Characteristics of Misinformation

The qualities that cause misinformation to go viral online are the same ones that make it so difficult to combat. In their study of rumor cascades on Twitter, Vosoughi et al. (2018) found that misinformation both travels faster and is more likely to be shared than true information, even when accounting for mediating factors like account age and number of followers. Through further analysis of the users' past posts, retweets, comments, and likes, the researchers postulate that the novelty and shock value of many misinformation posts often appeal to humanity's inherent attention towards new information (Itti & Baldi, 2009) and emotionally-evocative content (Berger & Milkman, 2012; Brady et

al, 2017; Crockett, 2017). This is exacerbated by the fact that the truth is often only presented in refutation of the misinformation, meaning that the false information has already had a head start.

Research has shown that the cognitive fluency of new information influences how easy it is to accept (Winkielman et al, 2012). This applies both ideologically (people more easily process messages that align with their beliefs) and semantically. For example, false statements like "smoke causes fire" (Fenker et al, 2005) and "Moses took two of each animal on the Ark" (Song & Schwarz, 2008) are commonly left unrefuted due to the associative nature of the content. Further, even more superficial characteristics have been shown to increase the believability of misinformation for audiences. Reber and Schwarz (1999) found that changing the text color to improve visibility and ease of processing increased feelings of familiarity with the text's content. Hearing information presented in a familiar accent (Lev-Ari & Keysar, 2010) or with rhyming within the statement (McGlone & Tofighbakhsh, 2000) were also shown to produce a similar effect. In contrast, participants in Song and Schwarz' Moses Illusion study mentioned above (2008) were more likely to catch the error (Noah was on the Ark, not Moses) when the statement was presented in a less legible font. This all suggests that people are more likely to believe misinformation when it allows for fluent mental processing, and that achieving such a level of processing fluency can be as simple as changing the font or adding a rhyme scheme.

In turn, misinformation is often intentionally designed to fabricate legitimacy (Di Domenico et al, 2021b). Fake news website domains and layouts are chosen to resemble those of legitimate news sites (Lazer et al, 2018), while headlines are written on the same relevant topics and utilize similar fonts and colors (Allcott & Gentzkow, 2017). Photos and images are digitally edited or misappropriated to seem relevant towards unrelated narratives (Shen et al, 2019; Tandoc et al, 2018). The recent proliferation of deepfakes, hyper-realistic videos and images developed through machine-learning, has introduced another layer of deceptive media into the misinformation landscape (Westerlund, 2019). As purveyors of

misinformation continue to co-opt common design heuristics as well as adopt new technologies, general audiences and researchers alike must stay vigilant to how these adaptations are being abused.

Internal Factors

Memory Effects

In its efforts to process incoming information efficiently through heuristic shortcuts, human cognition inadvertently makes itself susceptible to misinformation. The *False Memory effect* (also referred to as the *misinformation effect*) occurs when one's memory of an event is distorted upon retrieval (Britt et al, 2019). For instance, if a car-crash witness is asked to recall the accident, how the question is phrased (e.g. "how fast was the car going when it *smashed* into the other car?" vs. "how fast was the car going when it *bumped* into the other car?") can affect the witness's memory of the event (Loftus & Zanni, 1975). The *Illusory Truth effect* shows that repeated exposure to information gives individuals greater confidence in that information being true (Britt et al, 2019; Fazio et al, 2015; Pennycook et al, 2018). The repetition of a statement implants it in memory and makes it easier to draw upon, even if prior knowledge should seemingly refute it. Further, the *Continued Influence effect* theorizes that false information still influences one's recollection of information, even if new, correct information has been encoded (Britt et al, 2019; Lewandowski et al, 2012; Lewandowski & Van der Linden, 2021). Considered together, it is clear how misinformation can take advantage of these cognitive processes to force through confusion and misunderstandings: repeated exposure to misleading information surrounding an event can solidify that false perspective into an individual's memory as truth, and once that "truth" takes root, corrections only plaster over it but are not embedded as a direct replacement.

Source Effects

The source of information holds significant influence over whether the audience believes that information to be true or false. In their literature review on the topic, Brinol and Petty (2009) list the mechanisms by which source variables persuade the audience:

1. Source variables can act as a heuristic, standing as easily accessed evidence of their own argument based on their perceived credibility.
2. They can produce a relevant argument separate from themselves.
3. They can influence the amount of motivation the audience feels towards thinking about the information.
4. They can bias the direction that the audience's thinking takes.
5. And they can influence the metacognitive analysis the audience conducts on their thoughts (particularly, how confident they are in their own assessment of the argument).

How effectively the source variable can influence these mechanisms greatly depends on the context of the situation, including the audience's investment in the information, how the variable presents the information (e.g. speed of speech), and the variable's perceived expertise, attractiveness, power, and status (Brinol & Petty, 2009; Nadarevic et al, 2020). Nonetheless, source credibility has proven to be a reliable indicator of increased persuasion (Pornpitakpan, 2004). In terms of misinformation, Bauer and Clemm von Hohenberg (2020) found that people are more likely to believe and share information received from real news sources over fake ones (although the difference is small). In the same study, they also found that individuals are more likely to believe a fake news source when that source has previously provided information that was congruent with their beliefs. Similarly, Traberg and Van der Linden (2022) found that statements made by sources that are politically congruent with the intended audience are more likely to be believed, even when the statements themselves are nonpartisan. These findings imply that real

sources are generally preferred, but ideologically congruent messaging can potentially build up trust towards fake ones.

Di Domenico et al. (2021a) take a different approach, examining source effects when the source is a close personal tie (e.g. family member or friend) rather than a news outlet. They first examine the impact of source priming by telling participants beforehand to focus on the source of a presented fake social media headline, ultimately finding that priming the source reduces both trust in the message and willingness to share the post. In a second study, they conduct the same test but prime the source as either a strong tie ("closest friend") or a weak tie ("acquaintance"). The results revealed that priming continued to reduce trust and willingness to share for the weak tie group, but not for the strong tie group. This goes against Torres et al. 's (2018) study on social network ties and verification behaviors, where they suggest that weaker ties correlate with a decreased willingness to verify information before sharing (although post-hoc tests negated the significance of this effect).

Motivated vs. Classical Reasoning

A popular theory implemented for explaining susceptibility is that of motivated reasoning, particularly towards political misinformation (Kahan, 2017; Van Bavel & Pereira, 2018). Proponents of this approach suggest that human judgments are largely shaped by their own personal beliefs, principles, and sense of morality (Ditto et al, 2009; Kunda, 1990). When applied to misinformation, this means that "someone starts out their reasoning process with a predetermined goal…so individuals interpret new (mis)information in service of reaching that goal" (Van der Linden, 2022). As Wells et al. (2009) describe, this "systematic distortion of empirical beliefs" (p. 956) can stem from a number of biased information retrieval processes, including but not limited to

1. *Confirmation bias*, where individuals seek out information that supports their currently held beliefs (Nickerson, 1998),
2. *Prior attitude effect,* where individuals rate information that supports their beliefs more highly (Taber & Lodge, 2006), and
3. *Disconfirmation bias,* where individuals place more scrutiny on arguments that do not support their beliefs (Taber et al, 2009)

At its most basic level, motivated reasoning is the idea that people evaluate the information they gather based on how it fits into their own world view. This leads it to being the theoretical basis for arguments regarding the relationship between misinformation and ideological congruence or, more specifically, political partisanship. People have been found to adopt beliefs that align with their political identity, even if they do not align with their personal ideology (Cohen, 2003). And as mentioned earlier when discussing source effects, audiences are more likely to trust news sources that align with their own beliefs, even if they know that source is fake (Bauer & Clemm von Hohenberg, 2020). In political discussions on social media, outgroup animosity was shown to produce higher levels of negative engagement (Rathje et al, 2021). A more controversial extension of the motivated reasoning argument is that greater analytic thinking abilities lead to greater polarization, as people with better numeracy skills and higher education levels have tested as more partisan on scientific debates like climate change (Kahan et al, 2012). However, replications fail to find the same results (Persson et al, 2021), and Linden et al. (2018) argue that the correlation between higher education levels and belief in climate change is likely better attributed to the scientific consensus on the topic rather than partisan bias.

Researchers like Gordon Pennycook and David Rand challenge the idea of motivated reasoning as the primary factor determining susceptibility, instead preferring the classical reasoning approach. This stems from the dual-process theory, where human cognition processes information in two ways: intuitively (quick, “gut-instinct” judgments) and analytically (deliberative judgments) (Pennycook et al,

2015). Their theory is supported by recent findings suggesting that greater analytic thinking abilities are positively associated with better truth discernment (Pennycook & Rand, 2019), as well as belief in science (Gervais, 2015), disbelief in religious and paranormal anecdotes (Gervais & Norenzayan, 2012; Pennycook et al, 2012), and rejection of conspiracy theories (Swami et al, 2014). They argue that "analytic thinking will positively predict the ability to discern between fake and real news, regardless of whether it is consistent or inconsistent with one's political ideology" (Pennycook & Rand, 2019, p. 40). Instead, Pennycook and Rand point to inattention as the culprit of susceptibility – that poor discernment of misinformation is largely due to a lack of careful consideration and, given more time and focus, people will improve their judgment accuracy, regardless of their partisan loyalties (Pennycook & Rand, 2021). Over the past few years, Pennycook has conducted several studies in defense of his assertions. In a 2018 study, he and his team showed that repeated exposure to misinformation impacted accuracy judgments regardless of whether the statements were politically congruent or discordant with the participants' beliefs (Pennycook et al, 2018). In a 2019 study, he and Rand reported that performance on a cognitive reasoning test correlated positively with accuracy discernment towards real news headlines and negatively with discernment towards fake headlines, with both factors unrelated to the headlines' alignment with participant ideology (Pennycook & Rand, 2019). They built on this finding in a 2020 study, where they again saw the same correlations between cognitive reasoning scores and accuracy discernment, even when controlling for source effects (Pennycook & Rand, 2020).

However, recent replication and reanalysis studies have challenged Pennycook's assertions. Ditto et al. (2021) conducted a secondary analysis of Pennycook's 2018 data and argued that the effect can only truly be attributed to reduced belief in politically congruent information (which returned to normal levels in follow-up tests), while belief in discordant information remained low at all stages. Batailler et al. (2022) reanalyzed Pennycook's 2019 data and found that participants were more likely to label headlines that went against their political views as false, with the headline's true accuracy having little impact. And Traberg and Van der Linden (2022) countered Pennycook's 2020 study by showing that source effects

greatly influence misinformation susceptibility towards politically-driven topics, as both conservatives and liberals were more likely to fall for misinformation when it came from a politically congruent media source. Participants deemed the sources aligned with their own beliefs as more credible, with liberals placing greater emphasis on credibility judgments than conservatives. What all of these results ultimately show is that both motivated reasoning and analytic thought likely influence one's susceptibility to misinformation, with future research needed to determine the extent of their impacts and how they operate in collaboration with each other.

Backfire Effect

The *Backfire Effect* stands as a point of contention in misinformation research (Swire-Thompson et al, 2020a). Based around the motivated reasoning approach, some studies have found that refuting false information can cause certain individuals to become more steadfast in their beliefs. Two versions of the effect have been proposed: the *worldview backfire effect*, where the individual chooses to defend their perspectives against conflicting information (Nyhan & Reifler, 2010); and the *familiarity backfire effect*, where the misinformation is repeated during the refutation and becomes further lodged in the individual's mind (Pluviano et al, 2017; Skurnik et al, 2007). However, recent attempts to replicate both iterations of the effect have failed (Guess & Coppock, 2020; Pennycook et al, 2018; Swire-Thompson et al, 2022), leading researchers to believe that it is, at best, an extremely context-dependent phenomenon that should not deter refutation-based interventions.

Emotions

Emotional state has been found to have some influence on how people process information. There are currently two schools of thought: the first theory, referred to as the *assimilative-accommodative model*, posits that positive feelings induce more intuition-based reasoning, while negative feelings trigger

more analytic reasoning (Bless et al, 2006; Forgas, 2019). Aligning this with Pennycook's work around classical reasoning and misinformation (while setting aside the aspects involving political partisanship), it can be thus argued that positive emotions reduce accuracy discernment while negative emotions improve it. The other theory, the *resource allocation model*, asserts that both positive and negative emotions increase the load on attentional resources, leading to increased reliance on intuitive reasoning that ultimately reduces discernment accuracy (Ellis & Ashbrook, 1988). A recent study by Martel et al. (2020) supports this approach, finding that increased reliance on emotion correlates with greater belief in misinformation, regardless of whether the emotions felt are positive or negative. Similarly, Li et al. (2022) reported that higher emotionality (both positive and negative) correlated with greater belief in and willingness to share Covid-19 information, whether true or not. They also found that negative emotions mediated the relationship between Covid-19 risk and accuracy discernment, although this effect could be negated by greater analytic thought.

Individual Differences & Demographics

Certain demographic characteristics and personal traits have consistently been associated with higher susceptibility to misinformation. Political conservatism and age are both considered major factors, as studies examining social media trends during the 2016 election (Allcott & Gentzkow, 2017) and general sharing habits on social media (Guess et al, 2019) both point to conservatives and older adults as being more impacted by misinformation than liberals and younger generations. However, old age has also been associated with lower susceptibility in regard to Covid-19 falsehoods (Roozenbeek et al, 2020), suggesting that older adults place greater focus on health matters in which they are at greater risk. Higher education status is also frequently associated with lower susceptibility (Scherer et al, 2021).

In terms of personality traits, Chen (2016) reported that neuroticism correlates with lower misinformation sharing rates while openness to experience correlates with higher sharing rates. Zhu et al.

(2010) found that high harm avoidance, cooperativeness, reward dependence, and self-directedness all also correlated with increased risk of false memories. And perhaps unsurprisingly, Lobato et al. (2014) reported high belief in paranormal activities, conspiracy theories, and pseudoscientific claims all positively correlate with one another, suggesting that belief in these types of "empirically unsubstantiated claim" ties into misinformation susceptibility.

Societal Trends

Federal and Corporate Interests

Federal officials and politicians have long been found to manipulate the information being disseminated among the public to achieve their political goals. The Bush Administration's aggressive messaging around WMDs in the Middle East stands as one of the most infamous modern examples, as they took advantage of the social, political, and media climate post-9/11 to draw public support for the Iraq War (Arsenault & Castells, 2006; Lewandowski & Van der Linden, 2021). Their tactics were so successful that large swaths of American citizens still believed the WMDs existed years after it was proven false (Jacobson, 2010). But some instances are more subtle: In the 1980s, the U.S. government began appointing panels of expert scientists to address and regulate human activities causing acid rain. As O'Connor and Weatherall (2019) describe, then-president Ronald Reagan used his influence to plant Fred Singer on the panel, a physicist with expertise outside the relevant sciences (p. 35-40). Singer continually argued against the consensus that had been reached by the rest of the panel, delaying and tampering with their findings until the watered down final report was ultimately used to justify federal inaction on the issue for the rest of Reagan's term. And while politicians can manipulate media and scientific narratives to shape public opinion and achieve their political agendas, they may also do so to simply bolster partisan loyalties. After his 2017 presidential inauguration, Donald Trump made a point of touting the event's crowd size as greater than that of his predecessor, Barack Obama (Hunt, 2017). A study by Schaffner and

Luks (2018) has found that politically active republicans dependably defend these claims, despite all available evidence showing the contrary. Similarly, Swire-Thompson et al. (2020) have studied the impacts of source effects in regard to politicians. Their results show that, although fact-checking false statements by preferred politicians does reduce participants' belief in those statements, it does not significantly impact their overall feelings towards the politicians (at least in the United States; Australians rated their politicians more negatively when shown a disproportionate amount of fake statements, but Americans did not) (Swire-Thompson et al, 2020b).These examples all highlight the myriad of ways in which politicians weaponize misinformation for political gain and how their authoritative status creates source effects and partisan bonds that are difficult for the truth to overcome.

Corporations have also taken to utilizing misinformation in order to pursue or protect their interests. For example, the tobacco industry spent decades fighting and sowing distrust against the scientific research confirming their product's negative impacts on health (Smith et al, 2011). This continues even today, as vape companies "downplay the risks and addictiveness of e-cigarette use…[misinformation] is in part responsible for the youth vaping epidemic of recent years" (Tan & Bigman, 2020, p. 281). Oil companies have taken similar measures regarding their product's role in climate change, suppressing scientific findings revealing oil's negative environmental impacts and even producing their own research with cherry-picked results that could create an illusion of debate on the topic (Grasso, 2020). This corporate-sponsored misinformation has broad reaching implications, affecting populations far greater than just those who consume their products.

Journalism and Media

Journalists and traditional mainstream media outlets serve as a primary source of information for large proportions of the population (Allen et al, 2020). Unfortunately, over the last few decades, federal

regulations maintaining ethical journalism standards expired while vast media empires took root, allowing for ideologically-biased news coverage to be spread on a national scale (Iyengar & Massey, 2019). Major media corporations frame stories depending on both their target audiences and their own interests (DellaVigna & Kaplan, 2007; Groeling, 2008). This has led to the current media landscape where biased and sensationalized news stories are par for the course, playing into (and bolstering) the partisan loyalties of their viewers (Levendusky, 2013).

When conducted in good faith, journalism operates under the principles of "balanced" and fair reporting of all perspectives (although how that balance is defined is fairly subjective) (Benham, 2020). This balanced approach, however, can cause greater public confusion when the perspectives being presented are unequal in substance. When scientists have reached a consensus on vaccines not causing autism (Clarke, 2008) or climate change being real (Boykoff & Boykoff, 2007), giving equal airtime to opposing viewpoints can provide them a false legitimacy or trick audiences into believing the truth to be unclear (Bruggemann & Engesser, 2017; Lewandowsky et al, 2012). This issue is made even more complex when accounting for the fact that journalists are 1.) expected to provide real-time updates on fluid situations where facts are often changing at a rapid pace, and 2.) often tasked with simplifying complex scientific findings they themselves might not fully understand and translating those findings to a general audience (Lubens, 2015). As Lubens (2015) writes in regard to the relationship between journalists and public health professionals,

> [Reporters] need to consider whether reporting too much detail about the science of the disease will overwhelm a public that really only wants to know how they will be impacted personally. On the other hand, reporters might consider whether declining to present all of the science in order to report on a level that the greatest number of people will comprehend will simplify the health problem so that it misleads the public or reinforces superstitions about the disease's causal chain and biological pathway (p. 2).

These factors ultimately place even the most well-intentioned journalists and media outlets at constant risk of incidentally spreading misinformation to the public.

The Internet: Bots and Algorithms

The creation of the internet has revolutionized the way people distribute and receive information, removing the traditional gate-keeping mechanisms like editors and instead allowing individuals to actively produce and distribute their own content (Lewandowski et al, 2012). Unfortunately, this has also opened the floodgates to the spread of misinformation from sources all across the globe onto social media (Allcott et al, 2019). Social bots have been programmed to impersonate real people and spread misinformation (Shao et al, 2017; Wang et al, 2018). Moreover, algorithms designed to match one's personal preferences guide them into echo chambers of like-minded individuals (Cinelli et al, 2021), where the lack of diverse opinions leaves a fertile environment for misinformation to fester (Tornberg, 2018). These new technological methods for spreading misinformation will only grow more advanced and realistic over time. Developing effective ways to identify and combat their influence is essential for maintaining a safe internet environment.

Overall, this section has examined the cognitive processes and heuristics that, in the mind's effort to work as efficiently as possible, inadvertently leave individuals susceptible to misinformation; the characteristics of misinformation's design that take advantage of those processes; and some of society's structural dynamics that make spreading misinformation both possible and worthwhile. However, given the countless ways in which people receive information about the world and the sheer complexity of the systems in which that information travels to reach them, many of the deeper and more intricate interactions within these systems have yet to be fully understood.

Current Interventions

There are a number of interventions that have been produced for counteracting the effects of misinformation. In this section, I will examine the commonalities and differences between these interventions, the theoretical basis supporting their approaches, and any benefits and limitations that are unique to each type.

Measuring the Impacts of Interventions

Before detailing the different interventions, it is important to understand how susceptibility is being measured to determine each interventions' effectiveness. To date, there is not a widely-used, validated instrument set as the standard for measuring susceptibility. Instead, the vast majority of studies in this field conduct their assessments using self-produced instruments all structured in a similar fashion: by curating a corpus of both true and false statements and asking participants to discern each statement's accuracy, afterwards aggregating their responses to determine the participant's overall discernment ability (Li et al, 2022; Roozenbeek et al, 2020). These misinformation statements are most often presented in the form of news headlines, due to headlines being a primary method of news curation on social media (Gabielkov et al, 2016). Certain studies intentionally frame their headlines to target potential moderating factors; one common example is for political partisanship, where researchers will include an equal number of conservative- and liberal-leaning headlines or headline sources to account for bias in either direction (Dias et al, 2020; Martel et al, 2020). A validated instrument has just recently been released in hopes of standardizing this process (Maertens et al, 2024), although its adoption by others has yet to be seen and its application to certain types of interventions (like games) appears limited.

Types of Corrections

Regardless of their theoretical basis, most interventions implement one of three types of "corrections" on misinformation in order to steer the misled individual towards the right outcome. Ecker et al. (2022) describe these as thus:

1. "A fact-based correction that directly addresses inaccuracies in the misinformation and provides accurate information".
2. "[Addressing] the logical fallacies common in some types of disinformation".
3. "[Undermining] the plausibility of the misinformation or the credibility of its source" (p.18).

Through these options, intervention designers can choose to target and attack the specific "evidence" supporting the misinformation, the purported logic guiding the narrative between that evidence, or the qualifications of the person producing the misinformation. All three approaches have proven effective at correcting misplaced beliefs in misinformation (Cook et al, 2017; Ecker et al, 2020; Hughes et al, 2014). Studies assessing the benefits of utilizing a combination of these approaches in a single intervention are limited, but the few that have been conducted have produced promising results (Tay et al, 2021).

Primary Theories

The current literature on misinformation interventions is organized around two prevailing theoretical approaches: debunking (fact-checking) and prebunking (inoculation theory). The primary difference between these two approaches is time of application: debunking is applied therapeutically after misinformation exposure has occurred, while prebunking is applied prophylactically to prepare and protect individuals from potential exposure. This section will examine the benefits and limitations posed by both theories.

Debunking (Fact-Checking)

Debunking serves as the standard approach used to counter misinformation. Also known as fact-checking, debunking involves refuting and correcting a falsehood after it has started to spread (Van der Linden, 2022). Its effectiveness at reducing misinformation belief has been tested across a broad range of topics (Chan et al, 2017; Walter & Murphy, 2018), including tobacco use (Smith et al, 2011), vaccinations (Yousuf et al, 2021), and climate change (Swire-Thompson et al, 2021). Further, these effects have been shown to stay active for at least 3-4 weeks past the intervention date (Carnahan et al, 2021; Kendeou et al, 2014). However, these results are fairly variable depending on the quality of the debunk. To this end, researchers have begun to develop an understanding of the best practices for effective debunking interventions (Lewandowski et al, 2020; Ecker et al, 2022). These include

1. presenting an alternative explanation, rather than simply retracting the misinformation (Chan et al, 2017),
2. clearly stating how the misinformation is incorrect (and reinforcing the correction to emphasize its salience) (Schwarz et al, 2016),
3. establishing the credibility of the correction's source (Vraga & Bode, 2018),
4. promoting any social norms and expert consensus that stand in support of the correction (Schultz et al, 2007; Van der Linden et al, 2018),
5. conducting the intervention using clear, simple language and graphics for easier comprehension (Van der Linden et al, 2014),
6. And providing identity affirmations to help guide the individual through any major disruptions to their worldview (Carnahan et al, 2018).

Even with these conditions met, it is important to remember that debunking will still only likely reduce reliance on misinformation but not fully eliminate it (Chan et al, 2017; Walter & Murphy, 2018).

Beyond traditional fact-checks, the automated intervention measures that social media platforms like Facebook and Twitter use are also generally debunking-based, including related articles suggestions (Lyons, 2017) and warning labels (Lima, 2020). Bode and Vraga (2015) examined Facebook's usage of algorithmically curated article suggestions as a means to correct posts previously flagged as misinformation. Participants were exposed to misinformation posts regarding GMOs and vaccines, along with two related articles that either offered corrections, confirmed the falsehood, or offered one of each. Their study reported positive results, with participants who previously held the misperception showing positive improvement (even across partisan lines), although they present the caveat that the algorithm may not always offer corrections in practice. Warning labels involve placing messages around disputed content to alert the user that the information might not be correct. While this saves the platform from having to entirely delete the content, several studies have found them to be ineffective in the long-term (Ecker et al, 2010; Grady et al, 2021; Pennycook et al, 2018). Pennycook et al. (2018) found that, with or without warning labels, prior exposure to misinformation led to an increase in perceived accuracy upon second viewing. Clayton et al. (2020) found that warning labels were more effective when they directly label news as "rated false" rather than "disputed", suggesting that firmer language is necessary to disrupt the influence of misinformation.

The major limiting factors for debunking-based interventions all tend to stem from the therapeutic approach. While rarely replicated, there is the slight chance for debunking to induce a backfire effect that ironically strengthens the individual's belief in the misinformation (Swire et al, 2020a). There is also the possibility that repeating the false claims induces the illusory truth effect and bolsters the information into their mind, although researchers assure that this, too, is unlikely (Lewandowski et al, 2020). An issue more likely to occur is the misinformation drowning out the correction through the continued influence effect (Lewandowski et al, 2012; Walter & Tukachinsky, 2020). And due to the sheer speed at which

misinformation travels, debunking each new case after it develops is simply not a viable strategy for stopping its spread.

Pre-Bunking (Inoculation Theory)

Recent misinformation intervention studies have focused on inoculation theory as a promising prophylactic alternative to debunking. Initially proposed by Mcguire (1961), inoculation theory is often explained in parallel to vaccinations: by both warning individuals of potential exposure to misinformation (often referred to as *motivational threat*) and providing them with protective "doses" of counterarguments before the event occurs (*preemptive refutations*), the individuals can be better protected from future attempts at leading them astray. "These messages are thought to do two main things: generate threat (or the realization that an existing, desirable position is at risk of being challenged) and motivate defensive protections, such as counterarguing against impending challenges" (Compton et al, 2021, p. 3). A recent resurgence of work on the topic has broadened out both the impacts and the potential applications of the theory. Mcguire's interpretation saw inoculation as primarily intended for reasserting cultural truisms that few people would argue against ("you should always brush your teeth"); however, recent studies have found success towards more contested topics (like climate change) where individuals begin at varying levels of agreement (Lewandowsky & Van Der Linden, 2021; Van der Linden et al, 2017). And while originally applied towards refuting topic-specific points of misinformation (*narrow spectrum)*, it has now been found that inoculating against the misinformation tactics being used more generally can provide broader protection against the technique, regardless of topic *(broad spectrum)* (Cook et al, 2017). Cook et al. (2017) were able to avoid polarizing conservative participants when inoculating them against climate change misinformation by only explaining the general technique used (fake experts) and only in the context of the tobacco industry, rather than discussing the specific evidence from climate science. Some researchers have further taken advantage of inoculation's broad spectrum capabilities by using it to introduce more skill-focused interventions, like news and media literacy training (Hameleers, 2022; Tully

et al, 2020) and accuracy nudges (Pennycook et al, 2020; Pennycook et al, 2021). By being able to counter misinformation techniques used across contexts and before a falsehood even spreads rather than correcting the specifics within each individual situation after exposure, inoculation represents a promising option for keeping up with misinformation's speed of distribution. That said, it does rely on some level of prediction regarding what new misinformation technique or topic is likely to break out next.

Many of the same fact-checking techniques used during debunking can also be applied preemptively. A recent study by Grady, Ditto, and Loftus (2021) examined Facebook's recent attempt at prebunking warning labels where the warning appears to the user before the information is shown. While groups who saw the inoculating labels initially showed reduced belief in misinformation when compared to groups who saw the labels at the same time, the rate of accuracy between both groups evened out after two weeks. These studies suggest that, while warning labels can be beneficial in the short-term, their impacts are limited if the same misinformation is seen again later on without the label attached. Clayton et al. (2020) also reported a spillover or "tainted truth" effect for general warnings presented prior to exposure to the fake headlines, where warning the participants led to reduced belief in the real headlines as well. This side effect has occurred in other inoculation studies (Micallef et al, 2021; Van der Linden et al, 2017), emphasizing a potential downside to inoculation methods.

Inoculation theory has commonly been implemented through what are considered "passive" means, where the participants have the refutations presented to them directly. Similar to debunking, this is often through verbal explanations or by providing reading materials to participants. However, there have been recent studies supporting the benefits of active inoculation, where participants are prompted to develop their own refutations (Roozenbeek & Van der Linden, 2019a; Roozenbeek & Van der Linden, 2020b). The most prevalent medium used thus far for active inoculation is videogames. These game-based interventions will be analyzed more deeply in the following section, but overall, they

represent a distinct new area of opportunity for inoculation-based approaches that debunking has yet to offer a close parallel to.

Little research has been conducted that directly compares debunking and prebunking, with the few studies currently published showing varying results depending on the topic and context (Bolsen & Druckman, 2015; Brashier et al, 2021; Jolley & Douglas, 2017; Vraga et al, 2020). For instance, Jolley and Douglas (2017) found prebunking significantly reduced belief in vaccine misinformation but debunking did not, while Brashier et al. (2021) reported that debunking improves accuracy discernment to a greater degree than prebunking. This could suggest that different techniques might be preferred in different situations, but at minimum it means that more studies need to be conducted. Tay et al. (2020) also points out a major limitation in the intervention studies currently available, and that is the strict adherence to *explicit misinformation*. Most of the interventions mentioned in this thesis, along with the measures used to assess their efficacy, are structured around explicit misinformation as the default without consideration for *implied misinformation* – that is, misleading information that is also not false. As has been shown, misinformation is rarely a clear determination of true vs. false, and is often influenced by a multitude of factors beyond what immediate evidence is available. Accounting for these other factors and considering each situation in context is important for the future development of effective interventions.

Game Interventions in Focus

As stated previously, games have proven effective as interventions in other fields of study, including health (Lu et al, 2013) and wellbeing (Pallavicini et al, 2018). And given that games serve as the default method for implementing an active inoculation intervention, analyzing those that currently exist for this purpose can help develop an understanding of their efficacy and reveal potential paths for growth. This section will consist of an in-depth analysis of the game titles that have been designed for misinformation intervention, as well as considerations for the viability of misinformation games as an effective intervention medium moving forward.

Search Methods and Inclusion Criteria

In order to focus this review onto misinformation game *interventions* as opposed to games that simply utilize social deception as a mechanic, all searches were conducted in academic databases to establish a standard of academic rigor. The following databases were used: *Google Scholar, Academic Search Complete, Web of Science, Scopus, JSTOR,* and *Proquest dissertations*. Certain additional game titles were also added through informal google searches; while these generally lack empirically tested designs, they felt worth including due to their development by prominent fact-checking organizations or research institutions.

Terms searched include *misinformation* OR *disinformation* OR *malinformation* OR *fake news* in combination with *game* OR *intervention*. This resulted in an initial total of 29 game titles. During abstract checks, the following exclusion criteria were applied. Work-in-progress games and misinformation games that do not include an explicit intervention-based objective were excluded. Studies documenting misinformation-detection training workshops that include small gamified activities were also removed. Finally, games designed to teach about topics that are often targeted by misinformation but without a significant focus on the misinformation aspect were excluded as well. For example, *CO2peration* (2020)

is a game designed to educate youth about climate science but while avoiding worldview biasing and misinformation discussion (Harker-Schuch et al, 2020). This game is therefore not included in this review. The final list for review consists of eighteen games (Table 1). Mirroring previous literature reviews examining game interventions (Eichenberg & Schott, 2017; Fleming et al, 2017; Pakarinen et al, 2017; Pine et al, 2020), I have organized my evaluations based on design characteristics, study methods, and outcomes. This is intended to place focus on the variety between the game designs, how they choose to approach misinformation intervention, and how effectively their design choices translate to reducing misinformation.

Table 1. List of the twenty misinformation game interventions under review.

Game Title	Developer	Year Published	Type	# of players	Target Demographics	Main Mechanic	Content/Narrative Type	Intended for classroom use?
Play Fake News	ISL	2017	Digital	Single player	Not Listed	Detection	Nonfiction/Nonfiction	No
Polititruth	Politifact	2017	Digital	Single player	Not Listed	Detection	Nonfiction/Nonfiction	No
Factitious	American University	2018	Digital	Single player	Not Listed	Detection	Nonfiction/Nonfiction	No
Fake it To Make It	Amanda Warner	2018	Digital	Single player	Not Listed	Creation	Fiction/Fiction	No
Newsfeed Defender	iCivics, Annenberg Public Policy Center	2018	Digital	Single player	High school and above	Detection	Fiction/Fiction	No
Bad News Game	DROG, Cambridge Social Decision-Making Lab	2019	Digital	Single player	15 and above	Creation	Fiction/Fiction	No
Lamboozled!	MASCLab, Teachers College, Columbia University	2019	Analog (Card)	Multiplayer	Middle school and above	Creation	Fiction/Fiction	Yes
The Fake News Game	DROG, Cambridge Social Decision-Making Lab	2019	Analog (Card)	Multiplayer	High school	Creation	Fiction/Fiction	Yes
MathE	Katsaounidou, A., Vrysis,	2019	Digital	Single Player	Not listed	Detection	Nonfiction/Nonfiction	Yes

	L., Kotsakis, R., Dimoulas, C., & Veglis, A.							
Harmony Square	DROG, Cambridge Social Decision-Making Lab	2020	Digital	Single Player	15 and above	Creation	Fiction	No
Cranky Uncle	John Cook, Monash University	2020	Digital	Single Player	Not Listed	Detection	Nonfiction/Fiction	Yes
FakeYou!	Clever, L., Assenmacher, D., Müller, K., Seiler, M. V., Riehle, D. M., Preuss, M., & Grimme, C.	2020	Digital	Multiplayer	Not Listed	Creation + Detection	User-generated	No
Fakey	Indiana University	2021	Digital	Single Player	Not Listed	Detection	Nonfiction/Nonfiction	No
GoViral!	DROG, Cambridge Social Decision-Making Lab	2021	Digital	Single Player	15 and above	Creation	Nonfiction/Fiction	No
Trustme!	Yang, S., Lee, J. W., Kim, H. J., Kang, M., Chong, E., & Kim, E. M.	2021	Digital	Single Player	Young adults	Detection	Fiction/Fiction	No
Missed Information	Aguilar, E. M., & De La Vega, L. S.	2021	Digital	Single Player	Not Listed	Detection	Fiction/Fiction	No
ChamberBreaker	Jeon, Y., Kim, B., Xiong, A., Lee, D., & Han, K.	2021	Digital	Single Player	18+ years old	Creation	Nonfiction/Fiction	No
Escape the Fake	Paraschivoiu, I., Buchner, J., Praxmarer, R., & Layer-Wagner, T.	2021	Digital (AR)	Single Player	12-18 years old	Puzzle (Escape Room)	Nonfiction/Fiction	No
Misinformation is Contagious	Barzilai, S., Mor-Hagani, S., Abed, F., Tal-Savir, D., Goldik, N. A., Talmon, I., & Davidow, O.	2023	Digital	Single Player	Middle school and above	Detection	Fiction/Fiction	Yes
Fabricating Reality: The AI Misinformation Challenge	Shi, W. (2023). Efficacy of Educational Misinformation Games. arXiv preprint arXiv:2305.09429.	2023	Digital	Single Player	Not Listed	Creation	Fiction/Fiction	No

Design Characteristics

Publication Year

All of the reviewed games were developed and published within the last seven years, with the earliest titles being released in 2017. This makes logical sense, given both misinformation's rapid rise into public awareness and the influx of research on the topic that followed. Coincidentally, the two oldest games on the list are *Play Fake News* (2017) and *Polititruth* (2017), the two non-academically produced inclusions (they were developed by tech firm ISL and fact-checking organization Politifact, respectively).

Gameplay Mechanics

Almost all of the games in this review place the player in one of two opposing roles: either as the detector of misinformation or as the producer. For detection, the mechanics vary based on narrative depth and whether the content being assessed is imaginary or pulled from real world headlines. For example, three of the games (*Play Fake News*, *Politifact*, and *Factitious*) operate like dating apps: the player, embodying themselves, are simply shown an actual post taken from a news headline or social media and asked to determine if it is real or fake. *Fakey* (2021) builds off of this basic design by adding "like", "share", and "flag" (fact-check) buttons to more closely mimic social media interfaces and increase the level of depth of decision making (MIcallef et al, 2021). *MathE* (2019) takes a different approach by showing the player full news articles (real or fake) and equips them with four verification tools: search engine, reverse image search, image verification assistant, and debunking site (Katsaounidou et al, 2019). This is intended to train players to verify the accuracy of information they are presented with online. *Newsfeed Defender* (2018), *Trustme!* (2021), *Missed Information* (2021), and *Misinformation is Contagious* (2023) implement fictional narratives where the player acts as a social media sleuth tasked with keeping their newsfeeds clear of misinformation (Bouygues, 2018; Aguilar & De La Vega, 2021;

Yang et al, 2021; Barzilai et al, 2023). The card game *Lamboozled!* (2019) enacts similar processes in both a multiplayer and analog format by having players formulate arguments towards a "context card" based on the "evidence" cards they draw from the deck (Chang et al, 2020). Self-producing both the narrative and content likely allows for greater control, as these games are able to integrate depth measures like fact-checking and explanations of misleading clues in ways that the others do not.

Of the seven misinformation creation games, six utilize incredibly similar structures. The core gameplay for *Bad News Game* (2019)*, Harmony Square* (2020)*, GoViral!* (2021)*, ChamberBreaker* (2021)*,* and *Fabricating Reality* (2023) involves the player taking on the role of a social media manager enlisted to produce misinformation for nefarious purposes. They are then guided through a series of text-based decision trees as they are taught about different misinformation techniques and their impacts. While *Bad News Game* provides a broad spectrum understanding of common misinformation tactics like impersonation, polarization, and conspiracy (Roozenbeek & Van der Linden, 2019a), *Harmony Square* frames its messaging around politics (Roozenbeek & Van der Linden, 2020b), *GoViral!* focuses on Covid-19 (Basol et al, 2021), and *ChamberBreaker* describes echo chamber formation (Jeon et al, 2021). Despite being analog rather than digital, *The Fake News Game* (2019) operates in a similar fashion, as players are assigned to roles dedicated to spreading different types of misinformation, including "the denier", "the alarmist", "the clickbait monger", and "the conspiracy theorist" (Roozenbeek & Van der Linden, 2019b).

The two games that stray from these design structures are *FakeYou!* (2020) and *Escape the Fake* (2021). *FakeYou!* takes the unique approach of adding both the creation and detection factors together into a multiplayer game, where players are asked to write fake headlines for the same images and then identify which headline is the real option (Clever et al, 2020). *Escape the Fake* introduces two novel elements, placing misinformation clues and quizzes within an escape room puzzle game that is played through an

augmented reality (AR) lens (Paraschivo et al, 2021). The player is tasked with identifying the truth and completing the quizzes to exit the room.

These last two games represent the possible directions that misinformation games can take. While the detection and creation models stand as proofs of concept in regard to whether serious misinformation intervention games can be made, they largely lack the creativity and depth necessary to account for the nuances and complexities found in modern misinformation situations. For example, only three of the games listed in this review (*Lamboozled!*, *The Fake News Game*, and *FakeYou!*) include multiplayer mechanics, when information sharing is inherently a two-person act. While *FakeYou!'s* mechanical depth is limited (at the time of review, judgments had to be made solely off of reading the headlines, without any other mechanics for investigation), it adds competitive and collaborative elements that the single player games lack. *Escape the Fake* breaks the mold in both platform and genre, but still reverts to quizzes for integrating misinformation training into the gameplay.

Target Audiences

Ten of the twenty games listed do not include mention of target audiences, neither in their game descriptions nor in their sample populations. Of those that do, all seem to set middle school as the earliest age for recommended play. *Escape the Fake* skews the youngest, defining their target audience as twelve to eighteen year olds (Paraschivo et al, 2021). The DROG-produced games (*Bad News Game*, *Harmony Square*, *GoViral!,* and *The Fake News Game*) standardize their recommended age range as either in high school or fifteen and older. Despite this focus on adolescent or young adult audiences, only four games (*Lamboozled!, The Fake News Game, Cranky Uncle*, and *Misinformation is Contagious*) express intention to be used in (and provide supplementary curriculum materials for) classrooms. Given that children are at a key stage of media literacy development and older adults are identified as frequent targets of misinformation, more designers should consider developing their interventions with those at-risk demographic groups in mind.

Study Methods & Outcomes

There is a notable amount of variety in the methods applied and measures used among the games. Given the relatively recent release of these game titles, it is understandable that the majority do not have large scale empirical studies yet backing their efficacy. Due to the lack of consistency in approach between studies, I will instead analyze the methods and outcomes for each game individually. *Play Fake News, Polititruth, Newsfeed Defender,* and *Cranky Uncle* had no empirical studies evaluating the effectiveness of their interventions and therefore are not included in the following analyses.

Factitious collected gameplay and demographic data over a three day span, with measures limited to accuracy, rate of decision making, age, education, and gender (Grace & Hone, 2019). Data analysis (n=45031 players who responded to the demographic survey) showed that older players had better accuracy while younger players had faster decision times. Higher education level resulted in greater accuracy, although this tapered off at the PhD level.

Fake It to Make It's study design involved a mixed methods design, including pre- and post-test media literacy assessments, gameplay recordings, interviews, and retrospective read-alouds to assess the game's effectiveness as a persuasive tool (Urban et al, 2018). Demographic questionnaires and the pre-test assessment were conducted prior to the assessment, with the post-test assessment, interview, and read-aloud occurring afterward. Based on the interview data (n=7), the researchers found that a greater emphasis on the persuasive aspects of the game (more immediate feedback, emphasis on in-game goals, more realistic elements, using Mayer's principles of multimedia cognitive learning) would make the game more useful towards its intended purpose.

Lamboozled!'s study process included playtest observations, teacher interviews, reflection worksheets, and student surveys (Chang et al, 2020). The findings (n=76 students) showed that the game

did produce transferrable results for the students and created an engaging learning experience overall. However, the teacher's conduct and implementation of the game into their curriculum heavily influenced the game's efficacy; those who prepared more and tailored the game into their media literacy teaching showed greater results, while those who prepared less or the game didn't align with their curriculum showed worse results. The fictional narrative, while providing a non-partisan situation for students to engage with, also added a "suspension of disbelief" that complicated both playability and the transferrable learning process.

Missed Information utilized a pre- and post-test survey design to evaluate players' fake news knowledge changes influenced by the intervention (Aguilar & De la Vega, 2021). Players (n=24) were asked both about their methods for identifying fake news and their confidence in their ability to do so. Overall, there was an average increase in confidence by .524 (on a 6-point likert scale). Players also expressed having learned an average of 3 additional identification skills after playing the game.

ChamberBreaker conducted a randomized controlled trial to test the efficacy of the intervention (Jeon et al, 2021). The study utilized a likert-based reliability measure that is comparable to the standard accuracy discernment tests commonly used. The results showed that players (n=882) improved their reliability assessments and have increased their general awareness of echo chambers.

Similarly, *TrustMe!* also used a randomized control trial to evaluate its effects (Yang et al, 2021). The researchers used McGrew et al.'s (2018) method for online reasoning assessment, where participants are asked to make accuracy judgments and then explain the processes behind their choices, as a measure for information discernment skills. The results (n=210) showed that participants in the game condition performed significantly better on this measure than those in the control group or those who were presented with the techniques present in the game but without the gamified elements.

Fakey was evaluated through a mixed-methods approach, including analytics on the collected gameplay data as well as a small interview study to better understand the player experience (Micallef et al, 2021). To measure the intervention's effects, the authors compared players' responses between mainstream and low-credibility content types, as well as controlling for the amount of rounds the player ultimately played of the game. The findings (n=8608 unique users) support that playing for three rounds or more correlates with better recognition of mainstream articles, but not low-credibility articles; however, the authors point out that low-credibility articles are not always fake, hinting at instances of implied misinformation increasing the depth of the mechanics and processing involved.

MathE conducted an online post-test user experience survey measuring approval of the gamification experience and the verification training aspects (Katsaounidou et al, 2019). Results were positive, as 66.4% of participants (n=111) approved of the news-verification training, 62.7% said it enhanced their digital literacy, 70.9% said it taught them that those verification tools existed, and 50.9% said it taught them to use the tools.

FakeYou!'s methods were limited to a small case study (n=53), with data collection consisting of gameplay data and a post-play questionnaire regarding the player experience (Clever et al, 2020). The only measures assessing the game's impacts as a potential intervention were self-reported items regarding how much the player felt they had improved over multiple rounds of play. Fifteen users ultimately reported discernment improvement over time, while eighteen users felt it raised their awareness toward false headlines.

Escape the Fake only included a post-test survey, including the short flow scale for determining player immersion as well as survey questions regarding story quality, willingness to play at home, knowledge improvement, and rating for the in-game interactive NPC (Paraschivoiu et al, 2021). While not a particularly indicative measure, players (n=49) still rated their knowledge improvement as low. While

this might imply that straying from the traditional design norms can reduce an intervention's impact, I would argue that it is better interpreted as an indicator to prioritize how the intervention's goals can be implemented within whatever design choices one decides to make.

Fabricating Reality only conducted usability testing (n=12), recording observed and verbal feedback during each participant's playthrough (Shi, 2023). This feedback was primarily targeted towards iterating on the design itself, with no mention in regards to its misinformation themes.

Misinformation is Contagious used a between-subjects experimental design, where participants were randomly assigned to a play or control group (n=84 for study 1, n=131 for test two) (Barzilai et al, 2023). The groups would either play the game intervention or an alternative language game, then complete a single post-test survey. This survey included assessments of their discernment accuracy toward covid-19 information, their sharing intentions towards those same statements, their personal beliefs regarding misinformation, and their self-reported knowledge of common information evaluation techniques. Two studies were completed using these same parameters, with slight adjustments being made to study 2's assessment to make the accuracy discernment items more diverse and challenging. The final results for both studies showed that the test groups produced better accuracy and sharing intentions than the control groups.

The Fake News Game tested its applicability through an initial pilot study where participants (n=95 high school age students) were randomly divided between the game condition and the control group. After completing the intervention phase (playing the game or watching a presentation), testing involved participants being asked to assess a fake news article. The measure used to evaluate their assessment was a survey with questions regarding how persuasive they found the article to be, as well as several open-ended questions to understand how the participant interpreted the information (e.g. "what is

the author trying to convince you of?"). The results showed greater generation of counter-arguments for the treatment group, supporting this method of inoculation approach.

Bad News Game stands as the most rigorously tested intervention in this list. Its initial test utilized an in-game survey where players (between n=14163 and n=14266, with variance on which levels players were currently at when assessed) were presented with six likert-based accuracy discernment questions before and after playing (Roozenbeek & Van der Linden, 2019a). Results showed small to moderate positive effects, although the authors point out that this is in line with the findings produced from most persuasion research (Walter & Murphy, 2018). Basol et al. (2020) then completed a followup to this study, this time implementing a randomized controlled trial format. Now using a randomized control trial design where surveys were not administered in the game environment, the researchers were able to expand the accuracy discernment measure to include 18 questions and added questions regarding confidence in their decision making. The results (n=197) supported the original study, showing the inoculation group to have both better discernment towards fake news items and higher confidence in that discernment than the control group. Building off of this, three additional experiments were then conducted to assess *Bad News Game's* effectiveness in the long-term (Maertens et al, 2020). Using a similar structure to Basol et al.'s (2020) study design, all three experiments added additional test sessions at varying delays after the initial pre-/post-test intervention session: experiment 1 tested one week after, four weeks after, and three months after (using the same test items initial session); experiment 2 tested two months after (same test items as the initial session); and experiment 3 only tested one week after (different test items from the initial test). The results for all three experiments showed a significant inoculation effect for the test group after the initial session (n=151, n=194, n=170, respectively). This effect showed minimal decay throughout each subsequent test session of experiment 1. In contrast, the inoculation effect did decay over the time between sessions in experiment 2, suggesting that the repeated testing confounded experiment 1's results. The results from experiment 3's one week followup session showed no significant difference from experiment 1's followup session, removing memorization of the

initial test items as a confounding factor for experiment 1's results. Thus, the authors found that *Bad News Game* can have potentially long-lasting effects for up to three months, although the confounding nature of repeated testing necessitates further research into the actual time frame.

The two offshoots of *Bad News Game* followed similar methodological standards. *Harmony Square* utilized a randomized controlled trial design, with players answering sixteen likert-based accuracy discernment questions before and after playing the intervention game, as well as confidence regarding their confidence in their choices (Roozenbeek & Van der Linden, 2020). The results (n=681) reveal significantly improved reliability judgments and confidence in the intervention group.
GoViral! evaluations followed similar structures to *Bad News Game*'s first two studies. Both conducted by Basol et al. (2021), the first study analyzes pre-/post-test survey data taken from players prompted in-game at the beginning and end of their play sessions. To account for the in-game environment, the measures used were limited to six accuracy discernment questions each test. The results (n=1771) for this first study show participants to have significant improvement in their judgments towards Covid-19 misinformation. The second study switched to a randomized controlled trial structure similar to the one used for both *Bad News Game* and *Harmony Square*, with the addition of an infographics condition (where participants are passively inoculated) along with the control group. The results (n=231) again showed significant effects in improving judgments of Covid-19 misinformation and confidence in those judgments, as well as reduced willingness to share misinformation for the intervention group.

Given how many games in this review have only been evaluated through initial case studies and player experience surveys, it is difficult to make any judgments regarding their effectiveness as intervention tools. For many of the titles that do attempt to assess their game's impacts against misinformation, the methods used, the measures applied, and the sample sizes collected all lack the validity and academic rigor necessary to take their results at full value. As such, more elaborate evaluation methods with larger sample sizes are necessary for building confidence in their effectiveness as

interventions. This further emphasizes the importance of developing a standardized measurement of misinformation susceptibility, as even the studies with rigorous methods like *Bad News Game* and *TrustMe!* might prove less effective if examined under a different set of measures.

The Future of Misinformation Game Interventions

I argue that games are a perfect medium for confronting misinformation. The many complex factors that influence one's susceptibility to misinformation can effectively be simulated into a fun virtual world. Moreover, the barriers that limit an intervention's impact – for instance, political partisanship – are more easily overcome when the designer has full control over the social parameters under which the intervention operates. This is also where active inoculation's broad spectrum capabilities become even more of an asset: a political conservative might be less receptive to watching a video on climate change but more open to playing a fun murder mystery game where they must learn to identify common misinformation techniques in order to catch the culprit. In situations where a person's misinformation belief is not a mistake but a point of ideology, being able to leverage games as vehicles of motivation and engagement might be the best way to guide them towards the right perspective. Many of the games described are designed to simulate a basic level of misinformation interactions – the player is shown information with little to no supplementary context provided, the player decides whether it is true or false with little to no pressure surrounding the decision. But as new games are produced that build off of these designs and add new dimensions that account for different areas of susceptibility, I believe games will become a primary intervention tool for combating misinformation's complexities.

Chapter 3: Designing a Misinformation Intervention Game Mod

As described in the previous chapter, many different game titles have been developed as possible interventions for misinformation over the past few years. These games reveal a number of prevalent design trends within the genre: most are single-player linear narrative games where players are positioned either as detectives trying to determine truth from fiction (detection), or as propagandists creating misinformation using common techniques in an effort to sow discord (creation). Detection games generally involve presenting players with a statement or headline and asking them to determine its validity, occasionally with methods for obtaining clues like artificial search engines. In most creation games, the player is guided through a series of text-based decision trees as they are taught about different misinformation techniques and their impacts.

In terms of research testing the impacts of these games as misinformation interventions, there is a notable amount of variety in the methods applied and measures used. The majority find promising results through short surveys and interviews based on initial playtests; however, given the relatively recent release of these game titles, very few have conducted large scale empirical studies to back their efficacy (Basol et al, 2020; Basol et al, 2021; Jeon et al, 2021; Roozenbeek & Van der Linden, 2020; Van der Linden, 2023) . The only game tested for its long-term efficacy showed consistent results for up to three months (Maertens et al, 2020). However, the authors note that repeated measures might be a confounding factor and that the frequent testing might have acted as a “booster shot”.

While the detection and creation gameplay formats are effective at teaching common techniques used to spread misinformation, they largely lack the creativity and depth necessary to account for the

nuances and complexities found in modern situations where deception arises. Single-player games with fictional narratives can teach the general techniques used to spread misinformation, but they cannot replicate the social pressures involved (even amongst strangers–see Kou & Gui, 2014) nor the constantly advancing techniques being introduced in real life. What's more, they lack the replayability (Frattesi et al, 2011) and motivation to play (Fuster et al, 2014) that collaborative multiplayer mechanics can provide (Wendel et al, 2010). Given the quick decay in efficacy found for many inoculation interventions, producing a fun, replayable game that can repeatedly bolster players' cognitive "immunity" to misinformation is the pivotal next step for this genre.

In this chapter, I use a design narrative (Hoadley, 2002; Klopfer & Squire, 2008) to detail the process of creating *DoomScroll*, a researcher-constructed modification to Innersloth's popular social deception title, *Among Us*. DoomScroll incorporates several new features with the goal of accentuating the misinformation aspects inherent in Among Us' base gameplay, introducing common tactics used by spreaders of misinformation on social media, and promoting effective discernment techniques used by professional fact-checkers online. Through the design narrative structure, I will explain in-depth the ideation and implementation decisions involved in the creation of this mod. In doing so, I hope to convey the potential benefits, nuances, and hurdles of modifying an online multiplayer "entertainment" game for educational purposes.

Phase 1: Determining the Design Goals

Targeted Game Elements: Social Interaction & Replayability

The educational benefits of videogames are well documented (Squire, 2003), with a growing body of research focusing on the advantages of multiplayer games specifically. Massive online communities like those within MMORPGs are built through collaboration and the transfer of information

between novice and veteran players (Steinkuehler, 2004). On the more competitive side, esports have been shown to improve communication and socio-emotional skills among its athletes (Steinkuehler et al, 2023). Even when paired with strangers, players learn to engage tactfully while proactively cultivating positive (if temporary) relationships with their peers (Kou & Gui, 2014). One primary reason for these impacts is that social interaction is often a prominent motivating factor for engaging players, particularly for educational (or "serious") games (Lepper & Malone, 2021). Both collaborative and competitive play have been found to improve the efficacy of serious games and increase player motivation (Cagiltay et al, 2015; Pareto et al, 2012), emphasizing the benefits of multiplayer design. Harteveld and Bekebrede (2011) have summarized what they see as the differences in individual vs. multiplayer approaches to serious game design. They suggest that individual games are preferable for data-centric, linear gameplay that promotes concretely defined learning objectives using formal rules set by the designers. On the other hand, they believe that multiplayer games are better suited for non-linear gameplay that conveys broader, more process-driven learning objectives through social rules developed by the players themselves as they interact.

Mustaro and Mendonça (2012) argue for the importance of replayability in serious game design, citing its impacts on player motivation. They point out that including multiple non-linear narratives can entice players to come back to the game after completion, as each experience will be different from the last. By promoting repeated play of the main game mechanics through adding variation to the surrounding narrative, serious game designers can ultimately enhance the educational aspects of the game by having players continue to practice the core skills being taught even after their first playthrough (Kücklich & Fellow, 2004). In their review of the currently available misinformation games, Grace and Liang (2023) also point out the benefits of repeated play, noting that "balancing efficacy often relies on a combination of narrative fiction and repeated engagement" (p. 4699). They suggest designers use episodic short-form narratives that repeatedly engage the players in efficient, emotionally provocative play.

Frattesi et al. (2011) suggest that there are five different factors that influence a game's replayability: difficulty, completion, social aspects, randomization, and the game's overall feeling or "experience". Escalating difficulty motivates players to practice and overcome the challenges presented. Indicators of completion (achievements, unlockables, loot systems) can draw players back, particularly if the game is designed to not be completed in a single playthrough. Both the rivalry of competition and the camaraderie of cooperative play can entice players to continually practice a game. Randomization can ensure that each player receives a novel experience upon subsequent playthroughs. "The experience" is largely subjective to the player and the different aspects of the game that appeal to them, making it more difficult to design towards.

For a game intended to cover an inherently social concept like misinformation, competitive multiplayer gameplay seems like the logical approach to produce process-driven learning. And given that past misinformation game interventions cite "booster" doses as the remedy for the decay of their effects over time, prioritizing replayability appears to be a reasonable way to motivate players to repeat their training.

Taxonomies & Characteristics of Misinformation

Thus far, there have been several attempts by researchers to develop taxonomies for the different types of misinformation produced (Jia, 2020). The only types of misinformation with a near consensus listing are satire or parody, which are generally described as the spread of disinformation following journalistic norms for comedic effect (Molina et al, 2021; Tandoc et al, 2018; Zannettou et al, 2019; Wardle, 2020). However, Wardle is the only one to note the risks of this content, warning that such messaging becomes less humorous and more dangerous as it spreads and new audiences become less aware of its original intent.

Wardle (2020) takes a granular approach, detailing specific instances of trickery that might occur within a single article rather than the types of articles overall. Each category progresses further in perceived threat levels. After satire (the least threatening), "false connection" refers to the clickbait often seen within headlines, using hyperbole and emotion to draw in readers. "Misleading content" is information framed to present a biased perspective; an example might be an author presenting statistical evidence that points in one direction but manipulating the graph to read in a different way for inattentive audiences. Similarly, "false context" involves taking true information and completely changing the context in which it was produced (for instance, using images of past protests and captioning the image with the current date). "Imposter content" consists of disinformation that has been branded with the images or names of more trustworthy sources. "Manipulated content" is content that has been directly altered, such as photoshopped images. And finally, "fabricated content" is information that is entirely fake, like fictitious claims or deepfakes.

Tandoc et al. (2018) use a broader framework by focusing on the overall intent behind the disinformation being produced, as well as its direct relation to journalism. "News fabrication" consists of any false connections, misleading content, false contexts, manipulated content, or fabricated content that serve the purpose of spreading false narratives in the form of traditional news. "Photo manipulation" is largely the same (Tandoc et al. refer to false context as "misappropriation" when applied to photos). "Advertising and public relations" covers fake news that serves a corporate purpose and propaganda is fake news that serves a political purpose.

Like Wardle, Zannettou et al. (2019) include fabricated content and clickbait (false connection). "Biased or one-sided" maps similarly to misleading content, focusing on stories written to a hyper-partisan perspective. Like Tandoc's team, they include propaganda for political disinformation. Zannettou et al. then differentiate by adding conspiracy theories, hoaxes, and rumors as separate categories. "Conspiracy theories" are stories specifically describing events with insidious plots occurring

behind the scenes, without providing any proof of such plots. “Hoaxes” differ from fabricated and misleading content in that they define fabrication as entirely fictional stories, whereas hoaxes are largely true stories containing falsehoods or inaccuracies. “Rumors” are stories that are never confirmed, therefore the truth behind them remains ambiguous.

Molina et al. (2021) determine their categories based on a series of journalistic standards, including message and linguistic quality, sources and intentions, structural features, and metadata indicators. “False News/Hoaxes” will include factors such as fabricated content, emotionally charged clickbait, biased reporting, unverified sources, and poor or disreputable aesthetics. “Commentary, opinion, and feature writing” can be factually correct and come from known sources, but will also often be biased and emotionally charged. “Misreporting” will come from known and often reputable sources that meet general journalistic norms, but have simply not fact-checked or not verified their sources for the information. “Polarizing or sensationalist content” is similar to commentary but more biased, often with less evidence, less-than-factual information received from one-sided sources, dramatic rhetoric including clickbait, and increased emotional provocation and sensationalism. “Citizen journalism” can vary in message validity, and so is better defined by its lack of journalistic standards and its structural/aesthetic differences from more official sources (e.g. shared on a blog or personal site rather than a news organization’s website).

West and Bergstrom (2021) detailed different characteristics of misinformation that are commonly seen in science-based misinformation but are largely applicable to mainstream and social media as well. There is a notable incentive for authors to produce hyperbolic or emotionally evocative clickbait headlines that attract higher numbers of readers. Similarly, stories (and studies) with a more dramatic narrative and impact are more likely to be published and shared. As mainstream and social media become primary sources for receiving scientific news, findings and statistics might be misrepresented or misinterpreted by individuals sharing the data with bad intentions or simply a weaker

understanding of the topic area The authors themselves might even make these errors, either accidentally or with the specific goal to mislead (Bergstrom & West, 2017).

Common Strategies for Identifying Misinformation

Much of the current literature on evaluating information recommends critically assessing the source from which you received it, the information itself, and the broader audience response to that information (Sharma et al, 2019); or, as Wardle and Derakhshan (2017) frame it, the "agent", the "message", and the "interpreter". When pertaining to web articles, fake or misleading domains are common indicators of untrustworthy sources. Both the direct source of the article and the sources being cited within the article are more likely to be falsified, embellished, or misinterpreted in fake news, so it is crucial to verify the legitimacy of both (Molina et al, 2021). The Contact/"About Us" section of a website will often provide insights into authors and their backgrounds (Agarwal & Bandeli, 2017). Zannettou et al. (2019) list many of the possible source types, including but not limited to bots, criminal/terrorist organizations, activist/political organizations, governments, journalists, and trolls. These source types can work individually or even in tandem, for instance with political organizations enlisting bots to spread their messaging at a large scale.

Critically assessing the content of the information is also key in determining its validity. In comparing the content differences between real and fake articles and headlines, Horne and Adali (2017) note that fake news articles are often shorter, less technical, less analytical, and often more redundant. Fake headlines are usually longer, utilize more fully-capitalized words, and often make more claims than real headlines. False or polarized content is also likely to be more emotionally charged (Molina et al, 2021). Past research has shown that individuals often trust news sites more than personal sites or other genres, but if they are unfamiliar with the site or its sponsors, they are likely to build their trust primarily by using design factors as key heuristics (Flanagin & Metzger, 2007). Disinformation that contains fake

statistics or fake evidence with fake citation is often perceived as more accurate, necessitating more critical review from audiences in order to overcome these tricks (Hameleers, 2022).

The patterns found in user responses are similar to those found for sources; Molina et al. (2021) even categorizes them within sources and identifies the environment as the third major factor (see sections *Background & Related Work: Characteristics of Misinformation* and *Societal Trends* for examples of environmental factors). However, in order to distinguish the producer of misinformation from those who simply receive and share it, it is worthwhile to consider users as a separate group. Zannettou et al. (2019) offers names for two such categories, titled "useful idiots" and "true believers"/conspiracy theorists (p. 6). Useful idiots refer to otherwise normal individuals who, out of manipulation or naivete, trust information while staying unaware of the source's goals. True believers differ in that they truly believe in the information they are sharing. Examining an individual's shares, "likes", and comments on an article or headline can provide clues towards their motivations or even expose their true beliefs. Likewise, confirmations or refutations from more informed readers can also be shared within the comment sections (Agarwal & Bandeli, 2017). In an experimental study by Sterrett et al. (2019), participants reported greater willingness to believe and recommend a news article that was shared by a known public figure, even when the article originally stemmed from a fictitious news site. This shows that trusting the sharer can be a strong heuristic for trusting the content, even if an individual does not know the source of the content itself.

However, the investigative technique used is arguably just as important as the points of focus within the investigation. A highly touted approach for efficiently evaluating these different factors is to "read laterally" (Wineberg & McGrew, 2017). Wineberg and McGrew compared the strategies used by professional fact-checkers, historians, and high school students when evaluating online articles to determine the speed and accuracy of their techniques. They found that fact-checkers were significantly more effective in their determinations by reading laterally: they would spend little time on the actual

article, instead searching the internet for background information on the author and website itself to gain an idea of their overall trustworthiness. In contrast, many of the students and historians would read vertically–looking over the article and web page themselves for flaws (e.g. misspellings) or content-based hints of legitimacy (e.g. clean site layout, professional-sounding organization names, ".org" domains, lack of banner advertisements, etc.). Wineberg and McGrew argue that these quality-based heuristics were once promoted by digital literacy campaigns as quick indicators of website legitimacy ten years ago, but are far less effective under modern circumstances:

> At a time when the Internet is characterized by polished web design, search engine optimization, and organizations vying to appear trustworthy, such guidelines create a false sense of security" (Wineberg & McGrew, 2017, p. 31).

As such, they have developed curricula based around their Civic Online Reasoning (COR) approach (McGrew & Breakstone, 2023). This is intended to push individuals to ask themselves: Who's behind the information (what are the source's qualifications and motivations)? What's the evidence (and is it sufficient to answer the question)? What do others say (is there more reliable information or more trustworthy sources elsewhere on the internet to verify this information)? Caulfield's (2017) SIFT model takes a similar approach. In it, he suggests that individuals **S**top to evaluate the information before reading and sharing, **I**nvestigate the source, **F**ind (corroborate) with trustworthy sources, and **T**race the information back to its original context.

Thus, it can be argued that an efficient investigator of information validity will look out for these quick heuristics located within the information itself, as well as crosscheck their findings using outside resources to better understand the context surrounding the information.

Phase 2: Ideation

Design Principles

The problem of misinformation is by and large a social phenomenon. It is created by individuals (often for nefarious purposes) and spread rapidly among the populus as audiences share the information with their own social networks. As such, developing a multiplayer intervention seemed to be the obvious first step–not just for the benefits of social gameplay, but to more accurately replicate the contexts in which misinformation occurs and the pressures it instills. And as past titles in this genre have repeatedly shown, the positive impacts of misinformation intervention games gradually decay over time, emphasizing the importance of replayability as a vital draw for players to take additional "doses" and bolster their discernment skills. From these two points, the core principles we designed towards were:

1. Competitive multiplayer gameplay, to ensure that players must judge the validity of questionable information under challenging social pressures that might mirror real world situations
2. "Evergreen" narratives, where each playthrough provides players with new problems to solve within novel and compelling storylines

The biggest hurdles for these principles were readily apparent. The first was a more technical issue, as developing a sufficient multiplayer infrastructure with limited resources would be difficult and produce new issues that do not usually apply to single-player games (e.g. developing PvP communication systems, setting content moderation policies and tools, acquiring server space, etc.). The second was a design issue: developing a system for infinite replayability would seemingly require a daunting amount of content creation. For instance, for a game intended to teach lateral reading, a moderately replayable content base might involve several unique narratives each with a fictitious web browser and search engine, several websites, and dozens of accompanying articles. We soon came to the conclusion that for

truly infinite replayability, the content would likely have to create itself; in other words, the information would all have to be player-generated. However, allowing for player-generated content while also guiding players to misinformation discernment outcomes would still require strict control over the gameplay (not to forget the aforementioned communication and content moderation issues that would need to be addressed).

Among Us, developed by Innersloth, is an online social deception party game where one group of players (crewmates) work together to escape space station environments while avoiding or identifying a smaller, anonymous group of enemy players (impostors). Crewmates can win by either completing their assigned mini-game tasks (e.g. shooting asteroids, connecting wires, or refueling the ship) or by identifying and voting out the hidden impostors within their ranks during emergency meetings. The impostors must eliminate crewmates by killing them before their tasks are completed, while also sowing confusion and discord during group meetings to prevent their role from being revealed. By spreading misinformation and deceiving their peers, impostors can help their cause by placing suspicion on innocent crewmates and tricking the group into voting them out. In this way, crewmates are enacting a level of misinformation discernment in order to parse through the falsehoods and figure out who really is trying to deceive them. They are forced to investigate the validity of other players' arguments using whatever evidence is available to them as clues to the full narrative. Oftentimes, the majority of players are not directly involved in, or knowledgeable of, the incident being discussed; they were completing their tasks in a different room while the imposter struck, leaving them with little information beyond what is shared by others. And while the game cannot perfectly replicate the conditions of innocently encountering misinformation online (each player inherently has a stake in every dispute), I would argue this is a worthwhile tradeoff for players feeling greater investment in figuring out the truth. Despite these relevant design components, however, the base version of Among Us does have its limitations as a misinformation intervention. It has very few evidence sources from which players can draw conclusions – in most matches, text chat is rarely used (and the chat logs clear after each round), leaving players to rely solely

on their memory and verbal arguments. After assessing these factors, we ultimately realized that modding Among Us to include more misinformation-focused features was the correct path forward.

The decision to modify Among Us rather than develop a new serious game title was made for several reasons. As stated, the game already possesses a form of gameplay that draws on players' misinformation assessment abilities, providing a solid platform to build upon while removing the need to reinvent similar mechanics with new assets. The game's existing multiplayer infrastructure would also save us from having to produce our own server network and accompanying online policies. The developing studio, Innersloth, has a welcoming mod policy and so there already existed a vibrant modding community with numerous resources to ease our entry into the space. And maybe most importantly, Among Us is still a popular IP with a consistent player base, giving us an audience to design towards and that we could feasibly recruit to participate in our later testing stages. Given our development team's small staff and limited resources, this seemed like an appropriate path to develop a prototype under a realistic timeline. And also given that educational mods of AAA games are extremely rare and have predominantly been focused in formal classroom settings (Graham, 2014; Al-Washmi et al, 2014), we saw a unique opportunity to explore the potential of such work when designed for informal play environments. This mod was titled *DoomScroll*, in reference to the act of habitually surfing through social media while reading bad news.

Figure 3. The original logo for DoomScroll.

Learning Objectives

With the core development structure in place, we began shaping the learning objectives for DoomScroll and how they might be operationalized into gameplay mechanics. As shown in chapter 2, there are numerous different ways that misinformation can be structured and different approaches designers can take. Past misinformation game titles stand as perfect examples of this variety: *Bad News Game* and its spinoffs offer 6 different core tactics, including impersonation, emotion, polarization, conspiracy, discrediting, and trolling (Table 2) (Roozenbeek & Van der Linden, 2019). Upon closer look, these tactics each incorporate different combinations of source effects, emotional provocation, misleading content, and hyperbolic headlines. They are implemented into a fictional, single-player narrative where players are tasked with creating misinformation online using each of the different tactics.

Table 2. Common misinformation tactics (examples taken from Roozenbeek & Van der Linden, 2019).

Tactic	**Description**
Impersonation	Impersonating online accounts/other individuals or organizations
Emotion	Producing provocative content to increase engagement by playing into the audience's emotions
Polarization	Amplifying existing tensions between different groups in society
Conspiracy	Producing alternative explanations for events, often accusing a malicious groups of instigating such events
Discredit	Attacking or delegitimizing the arguments or credibility of opposing individuals or groups
Trolling	Inciting a reaction from target audiences

In contrast, *Fakey!* takes a different approach by having players evaluate headlines based around real world individuals and events from actual mainstream and low-credibility media outlets. Modeling the interface to those of platforms like facebook and twitter, the game allows players to investigate the headlines using Share, Like, Fact Check, and Skip buttons. The researchers argue that "Fakey! helps players learn to pay attention to various signals, such as clickbait headlines, partisan or emotionally-charged language and images, and context" (Micallef et al, 2021, p. 66). While both games are drawing on inoculation theory by teaching players common tactics and characteristics of misinformation before potential exposure occurs, the gameplay loops in which they implement the lessons are completely different.

Based on the literature as well as the previous games released, it would seem prudent then to include two pathways for improving discernment: teaching players common heuristics for determining whether information is credible enough, and teaching players effective strategies for investigating content further when those common heuristics are inconclusive. To more concretely lay these pathways out into actionable learning objectives:

- The player will learn common indicators of fake news, including questionable source domains, photo manipulation, emotionally charged or hyperbolic content, etc.
- The player will learn to investigate content using effective methods, including lateral reading and source evaluation.

Integrating the design principles (multiplayer and replayability) into these objectives would allow the players to practice these skills under realistic contexts and social pressures. If effective, the player might apply the knowledge and strategies gained from their play session when faced with dubious information in everyday life with greater confidence.

Phase 3: Development

The Challenges of Modding

We encountered two issues that complicated DoomScroll's development process. The first involved the technical challenges inherent to modding another developer's finished and compiled code base. Our programmers spent the first few months preparing the base plugin and familiarizing themselves with Innersloth's code base through the assembly code provided by the modding community. We identified the most relevant access points on which we could hook our own code so that the modded mechanics would be called during the appropriate gameplay events. However, the drawbacks of this approach quickly became apparent. While Among Us was developed using the Unity engine, our programmers did not have the luxury of utilizing Unity's visual editor, as games built in the Unity engine can only be modified through coded plugins and third party packages. As an example, rather than simply dragging and dropping each new button and asset onto the interface, we were required to estimate its intended size and position in code, enter a full match just to check whether it landed in the appropriate spot in-game, and then quit out and resize and reposition it in the correct direction. This slowed production significantly as small processes that would normally be achieved in seconds under standard development protocols now took from several minutes to several hours.

The second issue was more difficult to circumvent and led to several notable delays. Given that Among Us is a live game with several major updates (as well as smaller patches) occurring randomly throughout the year, we found that we needed to work quickly within three to four month windows to avoid having our progress disrupted by new features or unexpected changes to the source code. An early example of this was an update that negatively impacted the popular modding API we had been utilizing to install our mechanics. This delayed development for several weeks while we waited for the API's developers to identify and correct the problem. To remedy this, our lead programmer then spent an

additional week implementing a base plugin that would remove our reliance on that outside resource. The most disruptive example would occur several months later, when a major update reduced the permitted packet size that could be transferred through Innersloth's servers. This change inhibited our ability to let players send screenshots through Among Us' native chat function and thus effectively nullified one of our key mechanics. This delayed production for two months as we tested various workarounds before finally accepting the unfortunate event and beginning work on an additional mechanic to replace it. Fortunately, this would later become the headlines feature, which would prove to be the most important component of our final iteration.

These hurdles greatly influenced the design decisions made. We focused on adding features that worked independently from Among Us' code base and minimized the number of patched methods implemented. In a similar vein, we limited our mechanics to transmitting smaller datasets that would hopefully avoid disconnects and crashes caused by the unauthorized data transfers. These design choices allowed us to quickly adapt to any errors caused by future updates.

Modified Components

Our established learning outcomes and developmental parameters ultimately shaped the design of the initial iteration. The modded features implemented in this build were:

- *Persistent chat logs* – In base Among Us, chat logs (typed recordings of player statements and voting records) are available during each meeting until the voting phase has concluded, then emptied for the next meeting. We altered the chat logs to remain present throughout each match in order to allow players to use past statements and votes as evidence for their arguments later in the game.

- *Task Sign-In Forms* – Crewmates were given the opportunity to complete an assigned task, then reassign its completion to another player. This sign-in sheet was available to view by all players during each meeting as potential evidence of their innocence.

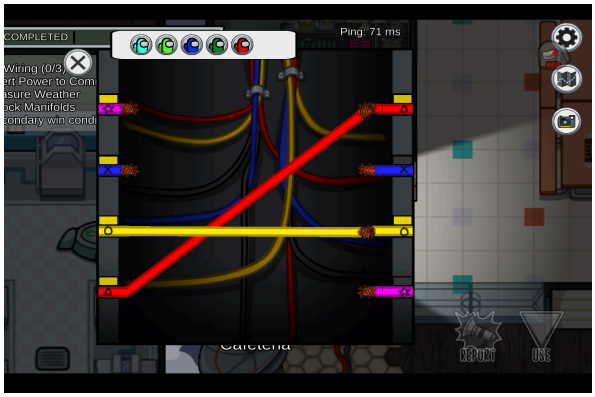

Figure 4. Task Sign-In function. Clicking one of the player avatars would list that player as having completed the task. This data would be listed in the folder system.

- *Headlines* – During task phases, a random player would be given the opportunity to "share" a headline and corresponding source from a randomized list. These headlines could be accurate clues based off of match data (e.g. "Player X has completed 3 tasks" and X actually has completed 3, indicating X is a crewmate) or falsified information meant to confuse ("Player X has completed 3 tasks" but X is an imposter and has completed 0). The sources acted in parallel, with some indicating trustworthiness (these sources always provided accurate information), some untrustworthy (these always shared false information), and some misleading or hyperbolic (these acted as "propaganda"). The shared headlines would later appear as potential evidence during meetings.

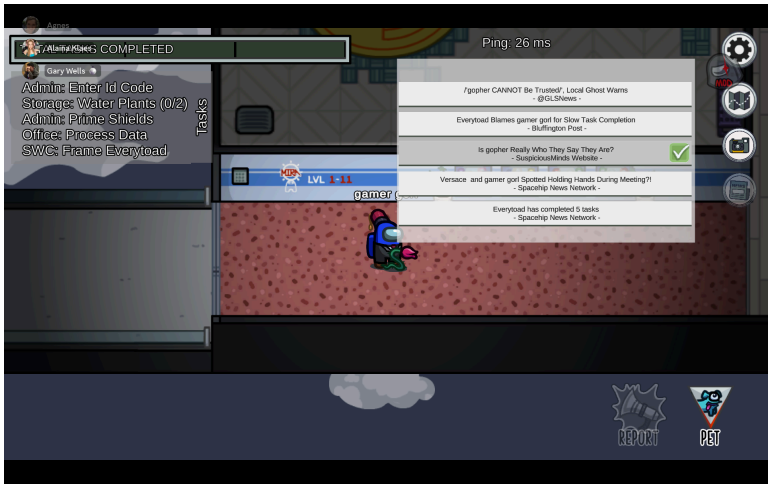

Figure 5. Headlines function. Randomly selected players could choose a headline to share with the group. Headlines were either true clues or falsehoods. False headlines contained common heuristics of misinformation.

- *Evidence Folders* – The folder system organized the potential evidence sources and made them available for all players to use in their arguments during each meeting phase.

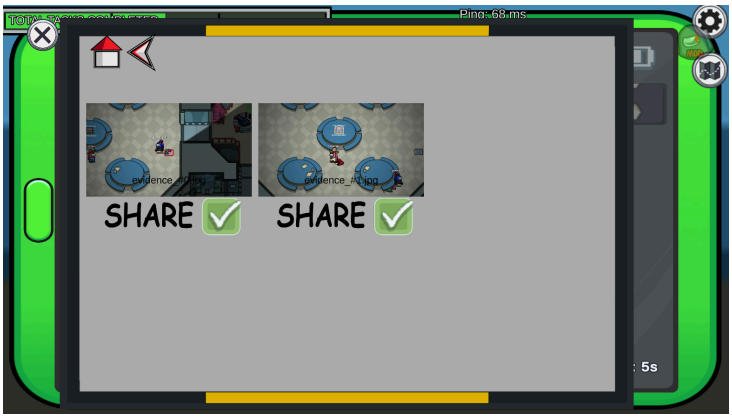

Figure 6. Evidence folders. Screenshots populated here and could be shared into the meeting chat.

- *Secondary Win Conditions (SWCs)* – SWC's added an additional objective for randomly selected players to complete as an individual. A player could be tasked with either "framing" or "protecting" another player, meaning they must try to get another player eliminated or keep them

from elimination by the end of the match, respectively. Players could complete these optional objectives in addition to their primary objective as either crewmate or imposter. While there was no in-game incentive rewarding players for completing their SWC, the final SWC results were made public on the final win screen. This was done with the goal of invoking social motivation between individuals by allowing them to "beat" or "align with" their assigned targets.

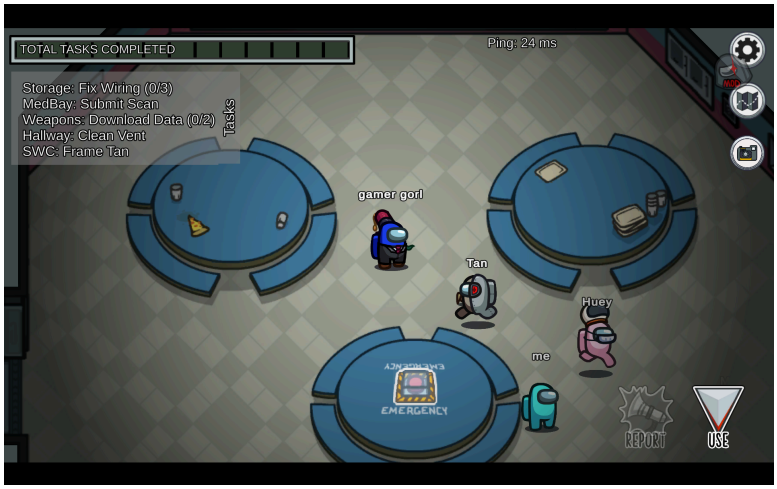

Figure 7. Secondary win condition (listed in task box). Players were tasked with framing or protecting another player. Success was determined by the target player living ("protect") or dying ("frame"). This was indicated in the task box as well as on the final end screen of each match.

The connections between the modded features and their intended learning objectives are detailed in table 3. Overall, the evidence-focused features were designed with the primary goal of improving the participants' ability to read laterally (Wineberg & McGrew, 2017). While developing a complex system of articles and search engine mechanics for this project was unrealistic, we implemented inherently flawed evidence sources that might prompt players to assess their value in relation to one another. One can read the fake headlines and simply search for obvious errors or indicators within the text and source names, or they can compare it to the other sources of information available. For instance, falsified task data can be fact-checked against the task sign-in sheet and in-person actions, or fake voting record claims can be reviewed in the chat logs.

We used the common content heuristics for identifying fake news as a guide for constructing all of the initial headlines and sources. Less trustworthy headlines were more emotionally evocative, made bolder (and more biased) claims, and used many of the same stylistic indicators (for instance, ALL CAPS). Source names contained subtle hints to a potential imposter bias ("The Red Report"), referenced less trustworthy site genres and sponsors ("Imposter Kyle's Personal Blog"), and possessed discrete misspellings of their more legitimate counterparts ("Spacehip News Network" vs. "Spaceship News Network"). Specific headlines contained more complex indicators; for instance, certain sources were listed as impartial databases that only reported match statistics (e.g. Task Archives could be paired with headlines reading "Player X has completed # tasks"). Both the trustworthy and untrustworthy sources could reference these database reports within their own headlines, potentially misrepresenting the statistics. In order to incorporate user responses and the complexity they add to the investigative process, each shared headline would indicate the player who posted it (in addition, an automatically generated headline was also produced each meeting). To decipher these headlines, the players needed to take into consideration the source, the content, the opposing player who shared it, and the current match context (including all other pieces of evidence) in order to determine its trustworthiness.

In real world situations, sources possess their own unique and complex motivations for spreading disinformation. Well-intentioned journalists will differ from politicians, who will differ from criminal organizations, who will differ from trolls. SWC's were intended to replicate this complexity by adding another layer of depth to each player's motivations and decision making. In base Among Us, a player can generally rely on their counterparts to be playing from one of three perspectives: a good-faith actor working to free the crew, a bad-faith actor trying to trick the crew, or occasionally a troll playing to create confusion and chaos. By adding SWC's, we hoped to foster more dynamic changes in each player's motivations over the course of the game. For instance, one crewmate might spend early meetings trying to frame another crewmate, only to then try and justify their treachery once that player is eliminated and

their focus must return to the primary objective. Or a crewmate might unknowingly protect an imposter early on, leading to suspicions against themselves as more evidence is revealed later.

The task sign-in form was designed to play off of the complexity added by the SWC's. Under normal Among Us conditions, seeing who completed which tasks would usually reveal who is and is not a crewmate (only crewmates are assigned tasks), incentivizing players to simply sign in as themselves. But when accounting for secondary objectives, players could now manipulate evidence in a way that might protect or frame an otherwise innocent target. This was intended to create situations where some players might argue for the validity of the sign-in data, while others might conspiratorially argue that it had been tampered with for nefarious purposes. Players could use this flawed evidence in tandem with other evidence sources to build a more stable argument. Overall, the task sign-in form was intended to invoke their investigative abilities and critical thinking skills, while potentially exposing them to different types of misinformation like conspiracy theories.

Table 3. The initial mod mechanics and their associated learning outcomes.

Mod Features	**Intended Purpose**	**Learning Objective**
Static chat log	To replicate social media by saving evidence for players to use in later rounds	Lateral reading – investigate source motivations
Task Sign-In Forms	Additional evidence source for arguments	Lateral reading – investigate information validity using available evidence sources
Headlines	Additional evidence source for	Lateral reading – investigate

	arguments	information validity by evaluating content; learn common heuristics of fake news
Secondary Win Conditions	Additional objective to influence player decisions	Lateral reading – investigate information validity by evaluating source credibility/motivations
Folder System	Organizational system for evidence sources	Lateral reading – investigate available evidence sources

Phase 4: Instructional Design Principles

After the initial features were completed, we analyzed the design decisions against standard instructional design principles to evaluate whether the core gameplay would operate in a way that could effectively push players toward its learning outcomes. Following in the footsteps of another multiplayer misinformation game intervention, *Lamboozled!* (Literat et al, 2021), we chose to evaluate DoomScroll's educational alignment with Gee's (2004) and Hobbs' (2010) instructional principles. Both were relevant to this project, as Gee focuses specifically on game design as a tool for creating engaging learning environments, while Hobbs focuses on the importance of media literacy in the digital age.

For Gee, DoomScroll aligns particularly well through the structured play environment it creates and through its integration of discernment skills into the core gameplay loop (Table 4). Each DoomScroll match places players into situations that simulate real world (or online) misinformation debates, while also simplifying the situations with limited variables and contexts to allow players to fully understand and

immerse themselves into the problems they are evaluating. The match objectives require players to learn and practice the same investigative strategies that they should apply when processing information in their everyday lives.

Table 4. Aligning DoomScroll with Gee's instructional design principles.

Gee's Principles (2004)	DoomScroll
Co-design: Good learning requires that learners feel like active agents (producers) not just passive recipients (consumers)	Players are given freedom to develop the narratives through their own creativity and gameplay
Fish Tank: In the real world, a fish tank can be a little simplified eco-system that clearly displays some critical variables and their interactions that are otherwise obscured in the highly complex eco-system in the real world. Using the term metaphorically, fish tanks are good for learning: if we create simplified systems, stressing a few key variables and their interactions, learners who would otherwise be overwhelmed by a complex system (e.g., Newton's Laws of Motion operating in the real world) get to see some basic relationships at work and take the first steps towards their eventual mastery of the real system	Real-world misinformation contains enormous amounts of contexts, variables, sources, etc. DoomScroll narrows this complexity down to a fictional narrative with simplified objectives and a handful of evidence sources. This makes it easier for players to practice assessing the information provided.

(e.g., they begin to know what to pay attention to).	
Sandbox: Sandboxes in the real world are safe havens for children that still look and feel like the real world. Using the term metaphorically, sandboxes are good for learning: if learners are put into a situation that feels like the real thing, but with risks and dangers greatly mitigated, they can learn well and still feel a sense of authenticity and accomplishment.	Similar to fish tanks, DoomScroll allows for players to practice their discernment skills within a contained, fictional narrative.
Skills as Strategies: There is a paradox involving skills: People don't like practicing skills out of context over and over again, since they find such skill practice meaningless, but, without lots of skill practice, they cannot really get any good at what they are trying to learn. People learn and practice skills best when they see a set of related skills as a strategy to accomplish goals they want to accomplish.	Gaining experience with common heuristics of false information and practicing one's discernment strategies are key paths to success in DoomScroll.
Cycles of Expertise: Expertise is formed in any area by repeated cycles of learners practicing skills until they are nearly automatic, then having those skills fail in ways that cause the learners to	DoomScroll provides a consistent gameplay loop with unique challenges presented during each match. This allows for players to practice their discernment skills under novel conditions until

have to think again and learn anew. Then they practice this new skill set to an automatic level of mastery only to see it, too, eventually be challenged.	mastery is achieved.

For Hobbs, the alignment between the principles and the game is much more clearly defined (Table 5). DoomScroll is designed to push players towards using the evaluation strategies and close analysis techniques that are vital to developing one's media literacy. The core gameplay allows for creativity and playfulness in one's arguments and decision-making while invoking discussion between players regarding the validity of different information sources.

Table 5. Aligning DoomScroll with Hobbs' instructional design principles.

Hobbs' Principles (2010)	DoomScroll
Using information search and evaluation strategies: Finding, evaluating and sharing content from a variety of sources helps people explore diverse sources of information. Using search strategies appropriate to one's needs helps people make discriminating choices about quality and relevance.	DoomScroll contains several different evidence sources in order to have players practice searching through them and evaluating their relevance to the situation.
Reading, viewing, listening and discussing: Active interpretation of texts helps people acquire new ideas, perspectives and knowledge and make	The core gameplay of Among Us (and in turn, DoomScroll) involves players actively discussing their perspectives and knowledge with others in

sense of it in relation to lived experience. Dialogue and sharing help deepen understanding and appreciation.	order to determine the truth.
Close Analysis: Careful examination of the constructed nature of particular texts encourages people to use critical questioning to examine the author's intent and issues of representation.	Critically analyzing the evidence sources is key to determining its validity.
Gaming, simulation and role-playing: Playful activities promote imagination, creativity and decision-making skills, supporting people's reflective thinking about choices and consequences.	The playful nature of DoomScroll's gameplay encourages players to both formulate creative arguments as well as make critical decisions that can win or lose them the game.

For both sets of principles, it must be said that the base Among Us game likely aligns with several of the principles through its own core gameplay. It creates a simplified but realistic information environment for players to practice discussing their perspectives and critically analyzing the information available, without any additional help from modded features. However, DoomScroll accentuates the skills necessary for misinformation discernment in ways that Among Us cannot currently reach on its own. For instance, while Among Us discussions are often reduced to verbal debates over what players remember seeing happen, DoomScroll's added mechanics provide the opportunity for players to strengthen their arguments by forming more complex narratives using the shared evidence sources. These sources are intended to amplify the need for investigation and critical analysis while also teaching players to clearly express their findings to skeptical audiences.

Conclusion

Based on the steps taken during this design process, we were confident that DoomScroll could work effectively as a misinformation discernment training intervention. Social and replayable gameplay elements can have significant positive impacts on players generally and, more specifically, on the efficacy of misinformation discernment games. DoomScroll was designed with these principles in mind. Its modded features promote investigative gameplay for players to seek out true information and discard falsehoods, while also incorporating different heuristics commonly used as touch points for identifying misinformation. With this prototype in place, the next section details phase four of this design narrative: playtesting. I will discuss the methods used for collecting player feedback and how adjustments were implemented in followup iterations to better improve the player experience.

Chapter 4: Playtesting a Misinformation Intervention Game Mod

Similar to its impacts on our design decisions, the risks of working with a live game significantly shaped our usertesting protocols. Given the short, three to four month windows between major developer updates (along with randomly occurring patches), we conducted the majority of user tests within two of these windows in an attempt to maintain continuity among the feedback collected. Working through a mod rather than a full game allowed us to focus solely on the added mechanics and how they interact with each other, both in terms of player experience as well as towards the overall learning objectives. Having smaller, text-based, complementary features–as opposed to larger, more complex, standalone ones–would hopefully allow for faster adjustments to be made based on the user feedback.

Methods

We ran nine playtests with ten different mod iterations over eight months, with between five to eight participants playing in each session and some individuals participating in multiple sessions (n=48). Players joined either in-person, online, or in a hybrid format, with the research team watching the online participants via shared screens on the Discord communication platform. Participants were recruited through convenience sampling of the research team's personal contacts and undergraduate volunteers from the host university. Our playtesting consisted of groups of five to eight participants playing between five and seven matches over the course of one hour, with researcher observations occurring throughout and focus group interviews taking place at the end. Players also completed pre- and post-test surveys containing both demographic items and three different assessments measuring their misinformation

discernment abilities in terms of truth vs. fiction, reliability, and civic online reasoning. The survey instruments were constructed and distributed using the Qualtrics software.

Participants

All participants fell between the ages of fifteen and thirty-two years old (M=25, SD=4.97). Participants' prior experience playing Among Us ranged from zero to one hundred hours of estimated playtime, with two participants reporting to have never played before (M=20, SD=26.86). Due to both the importance of written and verbal communication to Among Us' core gameplay as well as the DoomScroll mod being developed primarily for english-speaking audiences, all participants were required to be fluent in English.

Procedure

Participants were recruited to play the game mod via invitations sent through the research team's personal social media channels. Prior to their scheduled session, interested respondents were given instructions to set up the mod (if joining remotely) as well as a link to the consent form and pre-test survey. They were also provided time to read and complete these materials at the beginning of the playtest session. Once completed, participants were then prompted to open the mod and begin playing the game for approximately one hour. The research team observed and took notes on the gameplay as it occurred, as well as held brief, informal discussions between each match to ask how the players felt about specific features and rules. After playing for the allotted hour, players were then asked to fill out the post-test survey. Once completed, the lead researcher would conduct a brief focus group interview to receive additional verbal feedback on the experience as a whole.

After each session, the research team analyzed the feedback and actively adjusted the mod to better meet the players' needs. New iterations were produced and updated depending on the frequency of

feedback received and on the complexity of adjustments needed. The playtesting phase was concluded once an iteration was determined to have sufficiently met the learning objectives and player expectations.

Measures

Misinformation Discernment Assessment (8 item)

To assess the participants' misinformation discernment abilities, participants were asked to complete the Misinformation Susceptibility Test (MIST-8) before and after playing at least one match of DoomScroll (full instrument in Appendix A). The MIST-8 is the first validated instrument for measuring an individual's discernment abilities (Maertens et al, 2024), developed in collaboration with the research team behind several of the other RCT-tested game interventions (Roozenbeek & Van Der Linden, 2019a; Basol et al, 2020). Items were selected by their fit within the "Verification done" framework: news **v**eracity discernment, **r**eal news detection ability, **f**ake news detection ability, **d**istrust, and **n**aivete. The MIST-8 version consists of 8 items, 4 of which are AI-generated fake news headlines and 4 of which are real headlines published by reliable news sources. To match the initial construction of these items as well as to simplify the testing procedure, all items were rated on a true/false basis. The full instrument can be found in Appendix B.

Civic Online Reasoning (COR) Assessment

Three components of the Civic Online Reasoning (COR) assessment were used to test participants' lateral reading abilities. The components used were titled "News on Twitter", "Claims on Social Media", and "Researching a Claim". These were chosen due to their relevance to DoomScroll's content matter and because they each represent a different section of COR's curriculum: "Who's behind the information?", "What's the evidence?", and "What do other sources say?", respectively. The first two components present participants with real world tweets and ask them to determine the validity of the content and their sources. The third component makes a claim about a historical figure, then tasks the

participants with conducting research online to determine the claim's validity. All three are free response items with 8-minute time limits to allow for the respondents to explain their reasoning. The assessment materials used can be found in Appendix C.

Perceived Reliability Assessment

The perceived reliability assessment asked players to rate the reliability of 20 different statements, or "posts". The unreliable posts were developed to mimic real social media messages but still be fictional, in order to avoid memory confounds and to better integrate the learned materials into the assessment. Four reliable posts were also included; these are authentic posts made by real organizations and famous individuals. These items are intended to assess whether participants have improved their misinformation discernment abilities or if they have become more skeptical overall. All posts were presented in the same format as the headlines used during the game. Participants were asked to rate these items on a 7-point scale ranging from 1 – "Not reliable at all" to 7 – "Very Reliable". For each reliability item, participants were also asked to rate their own confidence in their judgment on a 7-point likert scale, from 1 – "Not confident at all" to 7 – "Very confident". Confidence is a proven indicator of one's conviction in their beliefs (Tormala & Petty, 2002) and an important factor in resisting persuasion (Compton & Pfau, 2005). Coupling improved judgments with an increased confidence in those judgments would be a strong indicator of DoomScroll's inoculation impacts. These items were included with the reliability assessment prompts. The perceived reliability instrument can be found in Appendix D.

Demographic Factors

After completing the intervention and the post-test misinformation assessment, participants were asked to complete an additional questionnaire measuring various demographic factors. These factors include age, gender, educational level, and political leaning. Research has shown there to be significant variance in how effectively different groups process misinformation compared to others (Allcott & Gentzkow, 2019). Further, prior studies using the MIST-20 (Roozenbeek et al, 2022) have reaffirmed the

theory that personal biases (often political) can influence one's misinformation discernment abilities (Van Bavel & Pereira, 2018). Additional questions were included to determine the participant's previous experience playing Among Us and Among Us mods.

Player Experience Measures

Players were asked to describe their play experience through several items on the post-test survey, as well as during informal discussions between matches and a focus group interview at the end of the session. The survey items included 5-item likert scale ratings regarding their usage of the various DoomScroll mechanics and whether they found the mechanics enjoyable to use. They were also provided free response prompts for players to describe any bugs they experienced during play, as well as any suggestions they might have to improve the play experience moving forward. The verbal discussions utilized the same items from the survey.

Analysis

After each playtest, three members of the research team would compare notes and observations to determine common themes taken from the player feedback and issues that had surfaced during play. Once a consensus was reached on the most prevalent points, we would discuss and implement new iterations that would potentially improve upon the previous build. Given the small sample sizes of each group and that each playtest was followed up by changes to the mod's design, the results from the survey assessments cannot provide significant findings regarding DoomScroll's efficacy. Given this, we instead conducted descriptive analysis on the responses, with the goal of providing both preliminary intervention data and to pilot test each assessment's viability to be used later in the full randomized control trial.

Findings

Almost every modded feature underwent several different iterations over the course of the eight playtests. This section will describe each feature's initial design, the feedback received from players, and the adjustments made in response to that feedback.

Task Sign-in Forms & Persistent Chat Logs

The sign-in forms and persistent chat logs were the only two features to stay relatively unchanged throughout the playtesting process. Given that the sign-in forms were fairly simple to understand and optional to use, players did not report any notable issues with the system and considered it fun and beneficial overall. The chat logs were only altered after the final playtest, where several players suffered frequent game crashes over the course of the session. The lead developer speculated that this might have been due to the larger group size producing a larger amount of messages that might have overloaded Innersloth's servers, although this was not verified. In order to prevent these crashes from occurring in the future, we limited the logs to only saving fifty messages maximum, after which the oldest messages would be deleted.

Screenshots

The screenshots feature was originally intended to be one of DoomScroll's primary evidence sources during its initial development, representing the manipulation and misappropriation of visual media (Shen et al, 2019). Every player would be allowed to take up to three screenshots of their avatar's immediate surroundings and share those images with their peers in the chat during meetings. Their personal avatars and user interface would be removed from the image to ensure that their roles and objectives remained secret. Unfortunately, unexpected updates from the developer limited the packet size that could be shared through their servers, hindering our ability to send images between players. After

several unsuccessful attempts to work around this issue, we temporarily abandoned this component to focus our efforts elsewhere.

By the final playtest, we had achieved a sufficient level of contentment with the other components and were able to return our attention to the screenshot feature. By revising the system to allow one player to send a single image per round (a new player is randomly chosen after every meeting) and breaking the byte array transmission into several packages with built-in delays, we were able to bring screenshots back to an operational state.

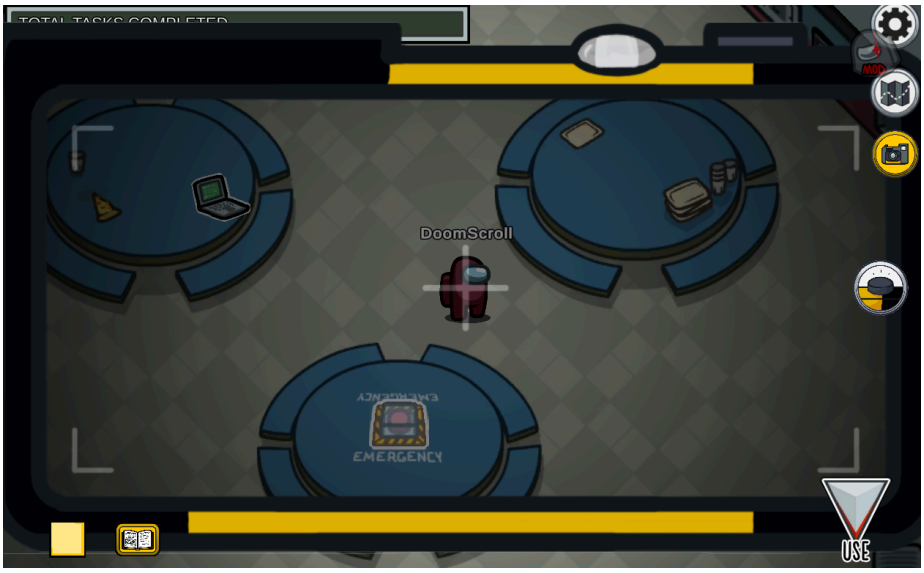

Figure 8. Screenshot overlay. By clicking the camera button in the top left, the player can open this overlay and save an image of their avatar's surrounding environment.

Secondary Objectives

SWC's were initially designed to act as a casual motivator for players to utilize the other evidence features. In the prototype iteration, some players would randomly be assigned a target player to frame or protect, with the results showing up on the final screen. We intended for this to encourage players to incorporate the new evidence into their arguments, such as signing in to tasks as their target to protect

them or arguing for news headlines that might frame them. Unfortunately, the SWC assignment text did not stand out to some participants, as it was only shown through a temporary role graphic that appeared at the start of every match. And even those who noticed it often forgot about their assignment as they progressed through the game. To ensure that players could remind themselves of their target, we added each player's SWC assignment into their task list on the main game HUD (head-up display).

However, these visual changes still did not motivate players to complete their SWC. While they did approve of its added complexity and expressed interest in it being added universally so that everyone would be assigned a target every match, researcher observations and player feedback confirmed that players were still solely prioritizing the main crewmate-impostor condition during meetings. During the post-session interviews, several participants noted that they had little reason to pursue their SWCs and suggested adding further incentives to more deeply integrate the feature into the game. To do this, we universally assigned every player an SWC and altered the requirements that needed to be met in order to trigger the final "Success" screen. If a player won the match as crew/impostor but failed to complete their SWC, the red "Failure" screen would now appear. This change effectively placed the SWC as another required objective to complete within the game loop.

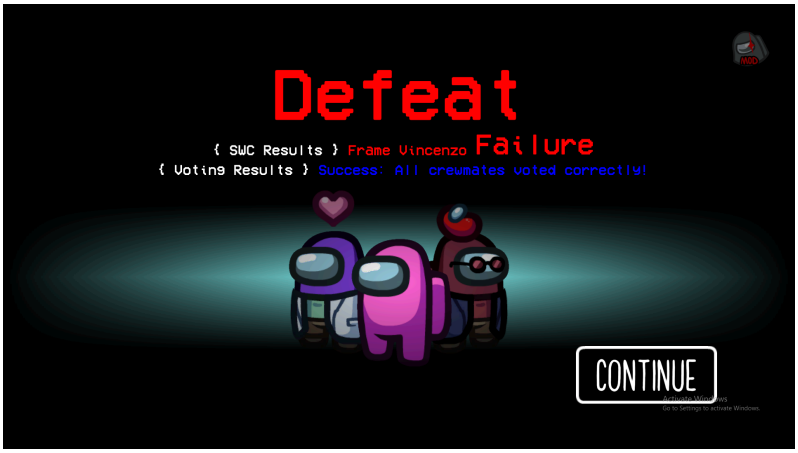

Figure 9. The final results screen. SWCs are now a required win condition; failure to complete them will result in the "Defeat" screen, regardless of if the crewmate-impostor condition is met.

Headlines

The headline system was initially intended to create an opportunity for players to produce misinformation and try to mislead teammates. At the same time, it would also require them to discern between true and false information posted by others. The first iteration allowed a random player to share one of five auto-generated headlines that could later be discussed during meetings. There were only 21 headline options available (12 trustworthy, 11 untrustworthy), with each post's trustworthiness being clearly labeled for the posting player.

One of this first iteration's issues was that the five allotted post options did not specify their frame/protect designations or targets, often making them irrelevant or even counterproductive to the person posting them. The feedback confirmed this, as players perceived the headlines to be humorous but optional content rather than an educational tool that might help them during play. Further, they felt that many of the post options were too obviously false to even share out to the rest of the group and suggested refining the content pool to make it more relevant.

From these comments, we restructured the headline pool by adding more options, including new posts as well as separating news and "user" sources. The new true headlines would have their information pulled from the match data, e.g. "Player X has completed # tasks" or "Player X has prevented sabotage". The news sources were more distinctly organized into trustworthy vs. untrustworthy groups, and user sources were introduced to impersonate biased individuals on social media (e.g. "Player X is completely untrustworthy!" -RedCrewmate84). The full headline and source list can be found in appendix G. To better integrate these posts into the gameplay, we assigned headline creation as a task to all crewmates and gave them the ability to select both the targeted player and whether to protect or frame them, resulting in three randomized options to choose from each time. Each post in the headline folder would be linked to the sharing player, with anonymous game-generated posts also being included each round. And to ensure

that players would have to not only post but read through and deeply analyze the headlines and source descriptions, we added an assessment-based win condition that the crewmates would need to complete in order to finish the game. After each meeting, players would be asked to vote on the trustworthiness of each headline (reliable vs. unreliable); to win, 75% of the crewmates (including members that had already been eliminated) were required to achieve 100% accuracy on these assessments, otherwise the game would continue (if the impostor was voted out, they would simply be revived back into the match). Finally, we added a "Source Background" folder into the folder system where players could investigate different content pieces providing insights into several of the news sources. These content pieces included a Wikipedia-like biography about an esteemed reporter, a news organization's homepage filled with trustworthy stories, an unreliable news organization's homepage mimicking its trustworthy counterpart's, and a personal blog site. These pieces each contain subtle references to the other news sources found in the game, giving the players the opportunity to practice lateral reading by critically analyzing the outside information provided and directing it towards their headline votes.

Figure 10. Headline folder redesign. Allows players to select a target and whether to protect or frame. The player can choose between three headline options (true/false is not revealed).

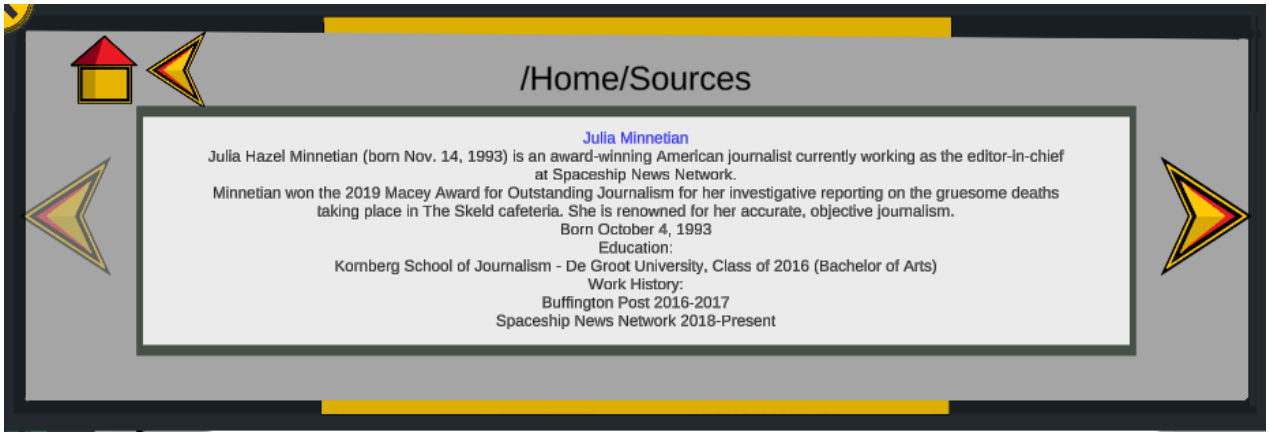

Figure 11. Source background folder. Provides players with context and insights into different headline sources and their trustworthiness.

These changes produced drastic improvements in players' utilization of the headline features. They showed increased engagement and more purposeful usage of the posts, including discussing their reliability with their peers during meetings. This improvement was made most obvious during meetings when the crew had already figured out the impostor's identity or had completed every other task, as they would spend several rounds solely focusing on each headline and the indicators potentially hinting at their reliability. One participant made clear that the increased emphasis on the new headline features caught their attention, pointing out in chat "we are apparently supposed to use the Sources tab to cross reference headlines" (S9P5). Another participant used events from the match to inoculate their peers against any potential falsehoods about themselves, proclaiming, "Snoopy and black [avatar] are confirming my innocence, (vote yes on those headlines later)" (S9P3). One player even later admitted to using the headlines to spread disinformation: "I just ended up putting a lot of fake news about [the lime green player] and I tried to spread a lot of misinformation about him too…But it didn't work" (S9P6). These adjustments proved to make the headlines system the most frequently used and likely the most important feature of the DoomScroll mod.

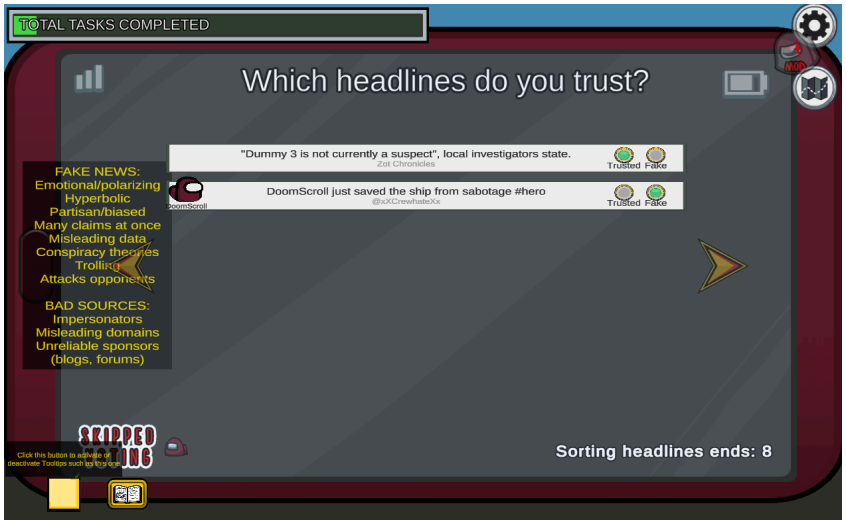

Figure 12. Headline voting. After the impostor vote, players have 20 seconds to vote whether each headline is reliable or unreliable. The supplementary tooltip box provides hints.

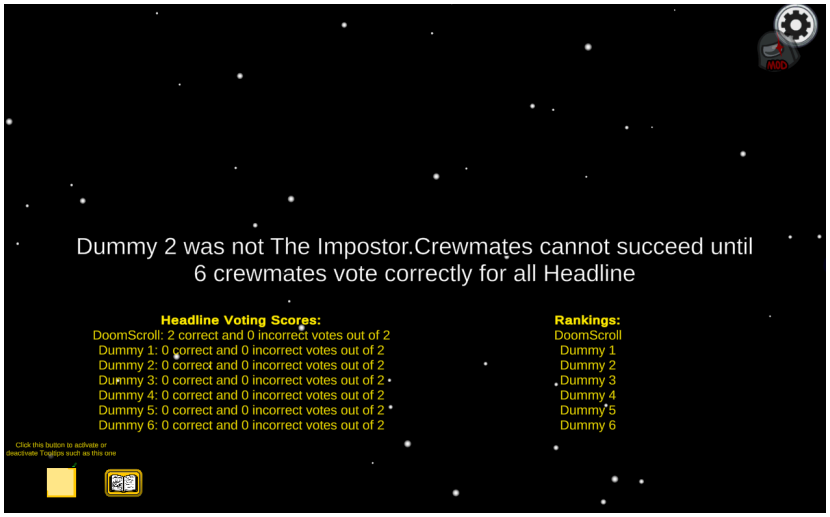

Figure 13. Headline vote results. After voting, players can see each other's accuracy on the previous headline assessment. They will be notified if they have achieved the accuracy % threshold to complete the game.

Help & Documentation

We had initially used Among Us' user interface, rule settings, and approach to instructions as a model for our own. Our assets were designed to match the existing ones in order to create an aesthetic cohesion between them. This led to confusion amongst players, as the lack of visual emphasis combined with the initial lack of incentives gave players little reason to click on these items during the run of play. We attempted to remedy this by highlighting these assets from the uniform grays to a bright shade of yellow. While this did improve their visibility, the changes in incentive and explanation were still necessary to increase their overall usage.

In the first iteration, players were expected to use Among Us' default settings; namely, having a minimum of 4-5 players, 15-second discussion and 120-second voting durations, and using voice chat to communicate with fellow players. After the first several playtests, the issues with these settings were made clear. Five players led to short games that prevented sufficient evidence from forming for them to use and analyze. The meeting length did not give players enough time to investigate, let alone discuss the little evidence that had accumulated. And when they would attempt to discuss the evidence, louder and more aggressive players were able to drown out the discussion by speaking over the others and dominating the conversation. To make sure that the proper settings were in place for the mod, we added a recommended ruleset to the waiting lobby HUD: 7 players minimum, 30 additional seconds onto the discussion timer, and they must exclusively use text chat to communicate. Followup playtests showed that this ruleset had the intended effects, as matches were long enough for evidence to be produced and discussed, the meetings allowed enough time for discussion, and text chat allowed for players to share the evidence clearly, in a manner similar to how misinformation is talked about on social media platforms.

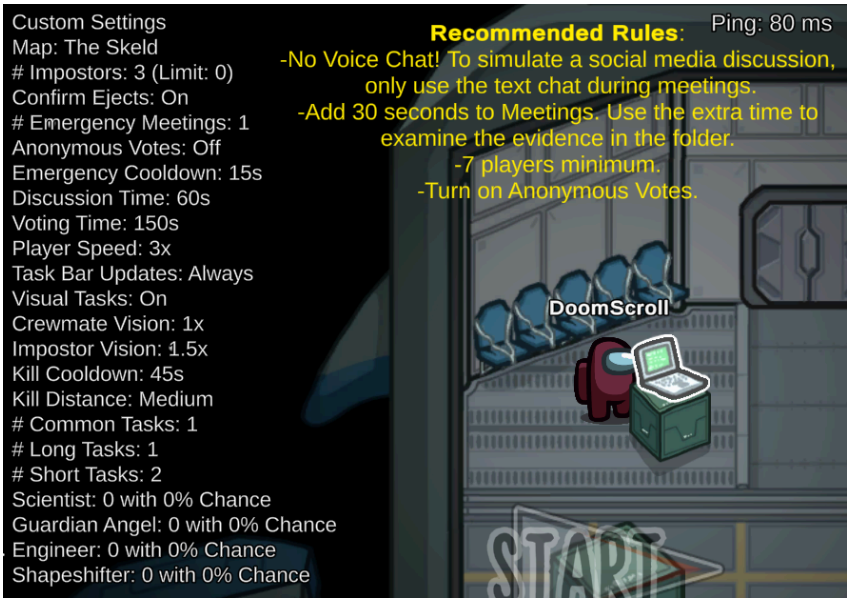

Figure 14. Recommended Rules. These rules encourage players to play DoomScroll in its intended form. This message appears in the pre-game lobby.

Similarly, we had attempted to copy Among Us' learn-by-doing approach by having minimal instructions beyond a short tutorial booklet available on the lobby HUD. Mirroring the interface issues, players were confused regarding what the modded features were and how they were intended to be used. It took them several matches to learn how to operate each feature in a meaningful way, leading them to instead focus on the Among Us components that they already understood. To give players more support, we made the tutorial booklet available to read throughout the match, in addition to in the pre-match lobby. We also added tooltips next to each modded feature that provide short explanations of each button's function, general hints regarding the common misinformation indicators used in the headlines, and reminders to practice their lateral reading abilities when investigating the evidence. Players were given the freedom to toggle these instructional materials on or off to prevent the amount of text on the screen possibly becoming overwhelming.

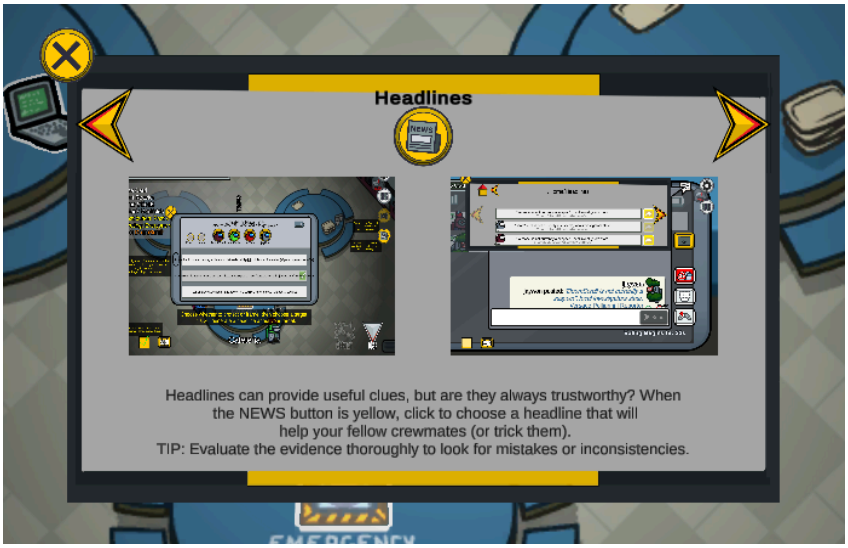

Figure 15. Instruction booklet. The instruction booklet provides descriptions of the mod's goals and features. Players can access the booklet in the pre-match lobby and at any point during the game**.**

Design Elements

The player feedback provided evidence suggesting that DoomScroll successfully implements our targeted game elements: social interaction and replayability. Across the nine different playtest sessions, 33 of 40 respondents reported interest in playing DoomScroll again in the future and only 3 said they would not play again. A common sentiment expressed was that the mod's complexity improved upon Among Us' base gameplay by introducing an additional layer of strategy and challenge, requiring greater collaboration between players. As Participant S3P4 summarized,

> "I find this mod more fun than vanilla Among Us, specifically because of the use of news headlines to gain additional insight about player relationships. I think this would be really fun to play with a large group, where the secondary objectives could really shine. Trying to get a teammate ejected while still trying to catch the imposter, in the end, is a challenging mechanic, which could be good or bad depending on the player. For me, it was a blast!" S3P4.

Pilot Notes

Each of the three test instruments were implemented into these playtests in part to determine if they would be suitable to utilize in the randomized control trial. The COR assessment notably took participants far longer than the others to complete, with many participants taking the entire 24 minutes allotted on both pre- and post-tests. Given that DoomScroll does not strictly adhere to the COR curriculum and to ensure that the RCT sessions could be completed in the expected 2-hour timeframe allowed by our host facility, I chose to remove the COR assessment from the study protocol. Instead, tooltip reminders were placed into the mod, prompting players to think about COR's three primary questions: "Who is behind this information?", "What is the evidence?", and "What do other sources say?". I chose to then assess DoomScroll's impact on their lateral reading abilities using qualitative means instead.

The ninth playtest group achieved promising scores on the reliability assessment, with five of eight participants improving from pre-test to post-test across the sixteen misleading items. It must be noted that one participant did score 13 points worse on the post-test; however, this score fell outside of the bounds indicated by the interquartile range (IQR=6.25) and so was removed from the dataset. This resulted in an average improvement of -1.57 (M=-1.57, SD=4.64). Players also showed modest improvement on the four real items (M=2.25, SD=1.39), suggesting that DoomScroll did not produce an unwanted increase in overall skepticism. Finally, most participants expressed increased confidence in their reliability judgments after the intervention (M=5.13, SD=7.12).

In terms of the MIST-8 Veracity metric (its primary measure of discernment ability), the results showed no notable improvement (M=0.13, SD=1.25). This mirrors the lack of significance found for this metric in the instrument's initial application study (Maertens et al, 2024), suggesting it might not be a suitable measure of discernment for this type of intervention. However, given that the MIST-8 is still the only validated instrument for this purpose and is short enough that its inclusion would not severely impact

the study's duration, both it and the reliability judgment assessment were carried over to the randomized control trial. With the intervention meeting its design goals and now appearing to produce promising results, we chose to conclude playtesting and proceed with the randomized control trial using this iteration of the mod.

Discussion

This chapter details the iterative process in which we adjusted DoomScroll's design based on playtest feedback. We were able to polish and expand on several of the features to increase player engagement while emphasizing the educational content. We introduced additional methods of support to help improve players' understanding of the project goals and to more smoothly ease them into the new gameplay loop. I was also able to use these playtests to both pilot the potential assessments that might be administered in the final evaluation study and to confirm that the mod appeared to have the desired effect on players' misinformation discernment abilities and their accompanying confidence in those skills.

Building from the design phase, these playtests provided clearer insights into the overall advantages and disadvantages inherent in modifying a mainstream game for educational purposes. For instance, the decision to mod did expedite certain phases of development by providing visual assets, the core gameplay mechanics, and a stable multiplayer infrastructure. Among Us' mass popularity also gave the mod immediate interest from audiences who could play the game with an inherent understanding of its rules.

However, we also experienced several drawbacks to modifying an existing title. Being able to build on a proven gameplay loop and utilize pre-made assets made the development process more efficient in some regards, but created barriers in terms of what we could alter and add to the game without access to the editor tools. And while having an audience of players already interested in the project was

helpful for recruiting experienced test participants, it might have given them pre-existing assumptions regarding how the game should be played. In a similar vein, we had attempted to match Among Us' design aesthetics and minimalist approach to instructions under the assumption that players would easily implement the modded features into their play, only for them to require far more instructional tools just to notice the changes at all. These factors all coincided during the playtests when participants instinctively reverted to ignoring the mod completely and focusing solely on Among Us' core objectives whenever they were confused. This suggests that a successful educational mod must integrate its material into the gameplay thoroughly enough to break the player's expectations, but it then also raises the question as to whether making a standalone title instead might ultimately be more effective. But despite these concerning trends in player engagement, the playtests had provided promising results, giving us confidence that DoomScroll was ready for the final evaluation.

Chapter 5: Evaluating a Misinformation Intervention Game Mod

Once the final DoomScroll iteration was complete, I conducted a more rigorous test of its effectiveness as a misinformation discernment intervention. This included examining the social interactions produced during play to assess the different techniques players use to investigate claims and determine their validity, particularly while other players are pushing to influence their beliefs. The research questions this study attempted to answer include:

RQ1. Does playing DoomScroll significantly improve one's ability to discern misinformation?

RQ1A. Do any demographic factors moderate this relationship?

RQ2. Does playing DoomScroll improve one's lateral reading abilities?

RQ3. What tactics do players use to succeed in the game? Are they effective, and how do they compare to the investigative tactics utilized by professional fact-checkers?

RQ3A. Do these tactics evolve as gameplay develops?

To answer these questions, I conducted a randomized control trial to compare the impacts of DoomScroll on misinformation discernment to those of a control group. In addition, I took exploratory observations of gameplay and player interactions to provide greater insights into players' decision-making process during the game. If successful, this study would support DoomScroll as a representative prototype for the future of misinformation inoculation game interventions. It would provide an attractive and effective training tool for preparing individuals to actively and accurately discern and avoid misinformation. Maybe most importantly, it would push game designers within the genre to produce more

impactful treatments by prioritizing replayability and socialization as important and necessary factors of misinformation intervention designs.

Table 6. The variables and data collection procedures pertaining to each research question.

Research Question	**Variables**	**Data Collection Procedure**
Does playing DoomScroll significantly improve one's ability to discern misinformation?	Misinformation susceptibility score (before and after intervention), reliability assessment score (before and after intervention)	Pre- and post-test assessment (MIST-8, reliability assessment)
Do any demographic factors moderate this relationship?	Age, gender, education level, political leaning, experience playing Among Us, experience with Among Us mods	Post-test survey completed after intervention
Does playing DoomScroll improve one's lateral reading abilities?	Source evaluation, evidence evaluation, investigation of claims	Gameplay observations, chat logs, telemetry data
What tactics do players use to succeed in the game?	Observed themes and patterns from gameplay	Gameplay observations, chat logs, telemetry data
Do these tactics evolve as gameplay develops?	Changes in observed themes and patterns over the course of gameplay	Gameplay recordings, chat logs, telemetry data

Methods

Test Conditions

Participants were randomly separated into one of two test conditions: playing the DoomScroll Among Us mod and playing *Tetris* (control). Tetris is an immensely popular digital puzzle game released in 1984 by designer Alexey Pajitnov. Players earn points by arranging and rotating falling shapes in order to form complete lines of blocks, which then clear away for new blocks to fall. Since its release, Tetris has been extensively researched both as a common control condition for studies examining more complex games and as the primary subject of studies assessing its potential as a cognitive training tool. It is an ideal control condition for this project due to its availability in the public domain, its simple and well-known ruleset, and its infinite duration (given that the game only ends once the player fails to clear enough rows and the screen is full of blocks, a single Tetris game can last as long as multiple Among Us matches). Further, it mirrors the protocol used by other researchers who have conducted randomized control trials to evaluate their misinformation game's efficacy (Basol et al, 2020). Using comparable media literacy literature as a control condition is also a proven approach (Jeon et al, 2021); however, given the length of gameplay necessary for collecting sufficient data on the test conditions, Tetris' versatile duration makes it a better match for this study.

Participants

Given that communication is an integral aspect of the game and to accommodate for developmental differences in misinformation discernment abilities (Loos & Nijenhuis, 2020) and media literacy levels (Arsenijević & Andevski, 2016), only english-speaking individuals ages 18 years and older were recruited to participate. To expedite the training process and to ensure that players were starting at similar skill levels, players were also expected to have prior experience playing Among Us in some capacity.

A power analysis was conducted in G* Power based on the results of previous studies (Basol et al, 2020). Using $\alpha = 0.05$, $f = 0.30$ ($d = 0.60$), a power of 0.90 and two test conditions, the analysis suggested that the target sample to detect the main effect should be approximately 120 participants. Accordingly, 127 participants were recruited for this study. 3 participants were removed for unreliable data, 1 was removed for incomplete data (missing pre-test), and 1 was removed for not meeting the eligibility requirements. The final sample consisted of 122 participants (n=122), with 62 in the control group and 60 in the experimental group.

All participants were recruited from the University of California, Irvine and were between 18 to 28 years old (M=19.79). A slight majority of participants identified as male (55%) and with liberal political views (53%).

Protocol

Participants were recruited to participate in the study through online posts, fliers, class announcements, and emails. Among Us requires between 5 to 15 players for each match and DoomScroll works best with 7 to 10 players. To efficiently meet these group size requirements, participant conditions were randomized by the session, rather than at the individual level. To accommodate different group sizes while also providing a standardized testing environment, all DoomScroll sessions and the majority of Tetris sessions were administered in UCI's Esports Arena. The arena provides uniform conditions and hardware for each player, while also allowing for enough room and isolation (there are enough stations that player seating can be arranged to prevent auditory or visual cheating) to reasonably replicate Among Us' intended remote-play environment. A small number of Tetris sessions were run in the Games+Learning+Society Center due to scheduling conflicts with the Esports Arena. However, the time constraints and technological parameters were kept nearly identical between the two locations.

At the time of their test session, participants were provided with information sheets and consent forms to read through and complete at their leisure. Once completed, they were then asked to complete a

series of pre-test assessments of their misinformation discernment abilities as well as provide additional demographic information. Once all assessments were completed, participants were asked to begin their assigned protocol. For the DoomScroll groups, this involved a short (15-20 minutes) introductory phase where the researchers explained the game's design goals and added features while each player followed along in-game on their own screen, testing each feature in practice mode. Tetris groups were offered an introductory explanation if they had no prior experience with the game. Both groups then played their assigned game for 1 hour. The playtests showed that DoomScroll groups could complete approximately 3 to 6 matches including the practice/tutorial (depending on group size) within this time frame, and that this amount appeared to be sufficient for producing the desired effects. During these matches, all players in the intervention conditions had their gameplay observed by the research team and textual communications recorded through the game itself. Control groups were asked to play Tetris for the same amount of time, but did not have their gameplay or communications recorded. Once their assigned protocol was complete, all participants were asked to complete the post-test misinformation discernment assessments. DoomScroll groups were then provided with a 10-15 minute debrief period where they were free to discuss their play experience and its perceived impact on their misinformation discernment abilities.

Measures

Misinformation Discernment Assessment (8 item)

This was the same assessment used during the playtests. The full appendix can be found in Appendix B.

Perceived Reliability Assessment

This was the same assessment used during the playtests. The full instrument can be found in Appendix D.

Attitudinal Certainty

This was the same assessment used during the playtests and was included as part of the perceived reliability instrument.

Demographic Factors

After the pre-test assessments and before the intervention, participants were also asked to complete a questionnaire measuring various demographic factors (Appendix E). These factors include age, gender, educational level, and political leaning. This data was used for exploratory analysis to determine if any of these factors acted as moderating variables towards the intervention's main effects. Players will also be asked whether they had any existing relationships with the other participants in their study session. Given the game's inherent social nature, this will provide insights into whether the intervention's efficacy might be moderated by playing with friends vs. playing with strangers.

Researcher Observations

Researchers observed players' DoomScroll gameplay throughout the study session, taking notes on any significant events or trends. There were at least two researchers in attendance at every single DoomScroll study session. Researchers compared notes and observations, then conducted a thematic analysis to determine the appropriate themes.

Chat Logs & Telemetry Data

Telemetry data collected from the DoomScroll sessions include chat logs, match duration, player roles, impostor kills, votes, meeting eliminations, evidence source usage, and final win conditions. This in-game data provided the research team with a clear transcript of each match, allowing for in-depth analysis of the events and how they might reflect players' information-processing and decision-making.

Debrief Interviews

Focus group interviews were conducted at the end of the DoomScroll sessions to debrief the participants and to gauge their overall experiences with the mod.

Analysis

Analysis for RQ1 involved using one-way ANOVAs to test the null hypothesis that the DoomScroll intervention does NOT produce change in an individual's misinformation discernment abilities or reliability judgments that is significantly different than the change produced by the control. Additional two-way ANOVAs were conducted to analyze RQ1a, testing the null hypothesis that the reported demographic factors do NOT significantly moderate the relationship between the variable groups and their measured impacts on misinformation discernment. All quantitative analyses were conducted in the R statistical software.

Analysis for RQ2 and RQ3 involved qualitatively coding the interviews, telemetry data, and researcher observations for common patterns and themes. The a priori codes used for this analysis can be found in Appendix F. In particular, the research team deductively analyzed the data for commonly used misinformation tactics and instances of lateral reading, as well as inductively analyzed for any other observed tactics that do not fit within the established categories. Two members of the research team independently coded one interview transcript and one chat log/telemetry data transcript, then met to discuss and adjust their codes until a finalized coding scheme was reached. The lead researcher then coded the rest of the data using that scheme. All qualitative analyses were conducted in the Dedoose analysis software. All identifying information has been anonymized using pseudonyms where necessary.

Results

RQ1. Does playing DoomScroll significantly improve one's ability to discern misinformation?

A One-Way ANOVA test was conducted to compare the effect of the experiment and control conditions on the difference in reliability judgment scores on misleading items between the pre- and post-tests. There was no significant difference shown between the tetris and DoomScroll conditions ($F(1,120)=0.06$, $p=.801$). These results indicate that DoomScroll did not provide a superior benefit on reliability judgment towards misleading content when compared to tetris.

A One-Way ANOVA was then conducted to compare the effect of the conditions on the change in reliability scores in regard to the real items. This test resulted in a weak but significant trend ($F(1,120)=4.14$, $p=.044$), suggesting that DoomScroll did improve participants' judgments towards the reliability of real content more than tetris.

A One-Way ANOVA test was also used to compare the effect of the treatment conditions on the difference in confidence scores from pre- to post-test across all items. Again, there was no significant difference between the conditions ($F(1,120)=1.7$, $p=.195$). While these results show a slightly stronger effect than the reliability judgment towards misleading content, DoomScroll ultimately does not produce a significant difference in the player's confidence in their judgment.

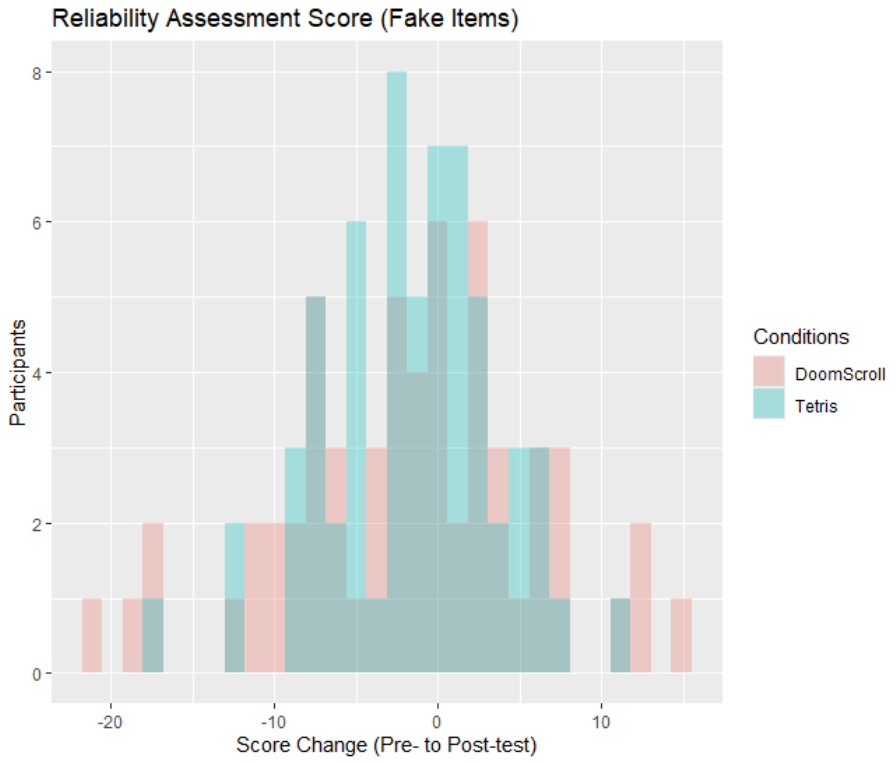

Figure 16. Histogram of participants' score change from pre-test to post-test on misleading items, by condition.

One-Way ANOVAs were conducted on the first three metrics of the MIST-8 *verification done* framework to compare both conditions' effects on the difference between pre- and post-test in terms of overall veracity, towards real news, and towards fake news . On the "Veracity Discernment" metric (measured as the sum of all correct responses), there was no significant difference between conditions ($F(1,120)=1.93$, $p=.17$). "Real News Detection" (the sum of all correct responses on true news items) was also insignificant ($F(1,120)=1.19$, $p=.28$), as was "Fake News Detection" (the sum of all correct responses on false news items) ($F(1,120)=0.99$, $p=.32$).

Two-way ANOVAs were conducted to test for moderating effects caused by demographic variables. None of the demographic variables produced any significant moderating effect on the

relationship between the intervention conditions and the shift in post-pre score differences on reliability judgments. See table 7 for details.

Table 7. Effects of demographic variables on the relationship between intervention condition and reliability judgements.

	Main Effect			Interaction		
Demographic Variable	F(,)	*F*	p-value	F(,)	*F*	p-value
Age	(1,120)	.032	.859	(6,120)	.933	.474
Gender	(1,120)	.063	.803	(3,120)	.199	.897
Education	(1,120)	.064	.801	(3,120)	.958	.415
Political Leaning	(1,120)	.066	.798	(4,120)	2.1	.086
Prior relationship with group	(1,120)	.064	.801	(1,120)	.662	.417

RQ2. Does playing DoomScroll improve one's lateral reading abilities?

Lateral reading is used by fact-checkers to "take their bearings" when confronted with new information. The civic online reasoning curriculum built around the lateral reading concept encourages

individuals to ask three questions when they are faced with new information: “Who’s behind the information?”, “What’s the evidence?”, and “What do others say?”.

In the DoomScroll chat, players frequently referenced each other’s potential motivations when discussing whose perspectives they trusted more strongly. The SWCs played a significant role in this (e.g. “[player 1] is trying to frame [player 2]”, “[player 3] is trying to frame me”). Some players would also search through the source background folder to acquire a better understanding of the fictional organizations and characters writing the headlines. As one individual described it during the post-session interviews, “You can discuss with the other people in the chat, but also looking through…the summaries of what the whole fake network is or whatnot. I guess that helps. It really ties them together basically” (S2P1).

Participants would also share and discuss the validity of different pieces of evidence. For instance, when one player shared a headline stating “*Art Professor: ‘As a doctor, I'm 100% certain that Timpo is NOT an impostor.’ -Innocent Street Journal*”, one of her peers sarcastically retorted, “art professors are always saying this”. Another participant reflected post-match on how talking about the evidence as a group was beneficial to their efforts: “I did see people utilize the headlines a lot or the task function. We mentioned, ‘oh, this person did X amount of tasks or this person hasn’t been checked in yet.’ I did see a lot of people utilize that and that definitely was helpful or contributed towards the conversation” (S3P6).

The following excerpt of dialogue from the chat logs shows a group of players addressing all three COR questions while attempting to complete their voting and headline objectives:

Mads shared: *SuperULT posted: “Is Viscidaide really who they say they are?” -@NotMe29*

Mads texted: yellow [SuperULT] trying to frame?

Liam texted: most likely

noodle has voted for SuperULT

Mads texted: also the only one to not do tasks

Liam texted: for headlines which ones do we think are fake

Mads has voted for SuperULT

noodle texted: im so confused with those no lie

Liam texted: 'noodle cant be trusted' gotta be fake

They assume that SuperULT has an ulterior motive to target another player and therefore disregard his previous assertions as disinformation ("Who's behind the information?"). They then utilize evidence from the task sign-in form to support their argument ("What's the evidence?"). By discussing their perspectives with the group, they are able to verify their beliefs, determine which other players might be trustworthy sources, and identify other headlines as false ("What do others say?"). While a more formal evaluation is required before any causal relationship can be claimed, examples like this suggest that DoomScroll does prompt players to practice their lateral reading skills.

RQ3. What tactics do players use to succeed in the game? Are they effective, and how do they compare to the investigative tactics utilized by professional fact-checkers? Do these tactics evolve as gameplay develops?

In every DoomScroll group, players began the session primarily collecting information regarding Among Us' main crewmate-impostor objective. The first message sent during the first meeting was often to ask "where was the body?" or to report "I was in [area of the map]". Some players attributed this in part to getting acclimated to the modded features, with one stating, "I think we all started out slow, sort of just getting the bearings of the game…I didn't know everything at first" (S3P1). Groups would only begin

to shift their focus once they had all finished their tasks or identified the impostor but failed the headline voting, forcing them to fully focus on the headline win condition in order to complete the match. They would then become more proactive about investigating the headlines early on: one player told their peers "guys i think we need to talk about the headlines more" right after a loss, while another reminded their group "we gotta vote for the correct headlines, [this player] is most likely imposter" to shift the focus to the next objective at hand.

As the matches went on during each session, players would become more comfortable integrating the evidence sources into their discussions. For instance, they would occasionally use the headlines and task sign-in form as an alibi to defend their innocence during meetings:

> Quiggles texted: I SAW JENKFEE COMING FROM THERE [where the body was found]
>
> Jenkfee texted: WOah i was in storage
>
> Quiggles texted: i watched u i just passed u
>
> Jenkfee texted: she's trying to frame me
>
> Amon texted: I saw cyan [Sevan] close to pink [the murdered player]
>
> Jenkfee texted: Sevan sus [suspicious]
>
> Sevan texted: i saw jenkfee near come from that place
>
> Amon shared: *Yee posted: "I've never seen Sevan kickflip even ONCE" -@JMinietian | Journalist @SpaceshipNewsNetw0rk*
>
> Amon shared: *Amon posted: "Tintedflag is a superhero sent to protect us we luv Tintedflag <3 <3 <3" -@Jenkfee_1515386*
>
> Amon texted: both fake right?
>
> *Jenkfee has voted for Sevan*
>
> Sevan texted: i can do a kickflip
>
> *Amon has voted for Sevan*

Quiggles texted: jenkfee sus, vote jenkfee

Quiggles has voted for Jenkfee

Jenkfee texted: u can literally see me finish my tasks

Sevan has voted for Jenkfee

Tintedflag has voted for Sevan

Quiggles texted: cmon Yee

Amon shared: *Jenkfee finished StartReactor*

Amon texted: btw

Sevan texted: that could be a lie

Yee has voted for no one

In this exchange, Quiggles accuses Jenkfee based on her eyewitness account of Jenkfee's location near the murder site. Jenkfee denies the claim, then takes advantage of Amon's suspicion of Sevan, who Jenkfee is required to frame. Sevan, the impostor, seconds Quiggles' suspicion of Jenkfee, then attempts to humorously defend himself by asserting that a false headline targeting him is true. Jenkfee defends her status as crewmate by pointing to the evidence that she has completed a task, with support from Amon. Sevan points out that the sign-in form could also be manipulated. Sevan is voted out, but survives because the headline accuracy condition is not met and ultimately wins the match.

While fact-checkers (hopefully) do not need to defend themselves from murder accusations, there are some parallels between their investigative techniques and those used by the players in this exchange. Jenkfee and Amon both correctly identify the impostor by discussing their personal accounts of the event and cross-referencing it with each other and with the available evidence. Tintedflag does not take part in the discussion but is also able to correctly vote out Sevan, likely in part by reading Jenkfee and Amon's discussion. Quiggles, who earnestly targets Jenkfee based on her own experience rather than an SWC,

votes incorrectly due to ignoring (or simply not trusting) the other players or the available evidence. Yee does not contribute to the conversation, nor does he cast his vote.

Jenkfee and Amon's investigative approach most closely aligns with the principles of lateral reading, and their success in this interaction supports its benefits. However, Tintedflag is able to achieve a similar level of success without actively contributing. His approach to passively engage was fairly common amongst the groups and extended to the discussions regarding which headlines were reliable and which were not. Here are excerpts from the same group as they attempted to complete the headline accuracy condition at the end of a different match:

Jenkfee texted: basically anything that doesn't sound like a news source is unreliable

Amon texted: any social media posts are unreliable

Yee shared: *"Sevan Saves Ship from Sabotage." -Amongus Weekly*

Sevan texted: is that the only fake one

Yee texted: this is real

Bobby texted: ok

Yee shared: *Sevan posted: "I can't believe anyone trusts Jenkfee" -@FreddyGazebo1922830485*

Yee texted: this also fake

Tintedflag texted: fake, yes

Quiggles texted: buffington, spaceship, among us weekly, jminnetian real. rest fake

Players would collaborate to quickly identify the correct answers to each headline. Those with a better understanding of the misinformation indicators would often simply state which headlines (or which sources in general) were reliable for the others to copy, rather than explaining the indicators themselves. Players who needed help utilized the "share to chat" component to point out any specific headlines that

they were particularly confused about. This suggests that clear communication and shared knowledge were popular and effective tactics for succeeding in DoomScroll.

Discussion

This study sought to determine the DoomScroll Among Us mod's efficacy as an intervention tool for improving player discernment towards unreliable and misleading information. The quantitative results ultimately suggest that the mod, in the examined iteration, did not produce any significant impact on players' discernment abilities towards misinformation when compared to the tetris control group. While its slight positive effect towards real items is encouraging, it raises more questions regarding how such a gap in judgment ability between the real and fake content could occur. Given the large difference in item count (4 real to 16 fake), it could simply be that adding in a comparable number of real items would decrease the variance of response between the two content types.

Looking back at the lessons learned during the playtests, the lack of effect could be attributed to the game's modded nature. Players might have focused more on the Among Us features that they already understood and less on the lessons taught by DoomScroll. However, the qualitative findings found through the in-game chat logs, researcher observations, and debrief interviews add complexity to these findings. The bulk of the chat text did primarily discuss Among Us' main crewmate–impostor objective: who was where, was anyone killed, did anyone see someone vent, etc. But over the course of the play session, several groups showed noted increases in the frequency and depth to which they discussed the different evidence sources, particularly the headlines. Players would share which sources to trust or distrust, occasionally referencing other pieces of evidence or the common tactics listed in the tooltips to support their beliefs. While DoomScroll does not provide a 1-to-1 equivalent to how lateral reading is enacted for traditional content found online, these findings still convey an improved process of

investigating information beyond its face value. What's more, this process is enacted in the nuanced social setting that we had initially set out to create. And so through this qualitative data, it becomes apparent that some degree of learning did occur.

This points to some of the key differences between a social game like DoomScroll and its single-player predecessors. The chat logs reveal that many sessions largely consisted of a portion of the players actively discussing the evidence, while others communicated only when asked or otherwise stayed completely quiet. This is par for the course for an average game of Among Us and reflects how players were given the freedom to play however they felt comfortable. One might predict, then, that some individuals engaged with the educational materials more than others, and that some chose to instead turn to their more informed peers as their own trusted sources. Thus, future iterations of this study might be better served by more closely accounting for players' active, passive, and absent engagement with the intervention (as well as length of exposure) when considering the overall "dosage" they each received. It must also be noted that the better informed players would not always explain *why* a headline was true or false, giving the answer but without the lesson learned. Such a trend aligns well with Harteveld & Bekebrede's explanation of the differences between single- and multiplayer games as educational tools:

> "Another reason to opt for a multiplayer game from a social learning perspective is…the common knowledge that a group of people are able to know and reach more than a single individual. On the other hand, in experiments, it is also been proven that groups show an average performance, that not everybody within a group contributes just as much, and that some people do not contribute at all (Forsyth, 2006). What if we were able to make all individuals the best people in class through a single-player game? We would certainly know who is slacking and who is not in that case." Harteveld & Bekebrede, 2011, pp. 52-53.

In this way, using a multiplayer game to add social elements might have produced unexpected negative impacts. Prior single-player misinformation games might not have the replayability or social interaction needed to draw long-term player interest, but they can ensure that every player achieves the intended educational benchmarks on every playthrough. DoomScroll lets players practice their discernment skills under realistic social conditions, but allowing for collaboration with others can lead to an uneven distribution of learning.

These findings hint at ways to improve on the testing procedures and the game design overall. By using Among Us as the base for this project, we understood that, if it were an effective and enjoyable intervention, it would most likely be played casually outside of traditional educational environments. As such, we had always designed the mod's educational features with the goal of them being complementary to the system as it exists, rather than overpowering the core gameplay loop and potentially negating the factors that players enjoy. For instance, we actively chose to include minimal explanations (e.g. the tooltips and instruction booklet) regarding the different types of spread techniques used and to have no results screen summary containing the correct answers for that match's headlines. I took a similar approach when designing the evaluation protocol, as players were given little to no instruction or explanation about the mod's goals or lessons and were allowed relatively free reign as to how they could play the game. Introducing more deeply integrated, explicitly educational features into future iterations of DoomScroll would likely improve its effectiveness as an intervention by forcing players to engage with the materials in order to succeed. Implementing the intervention protocol with additional instructional features (Pilegard & Mayer, 2016) and more defined learning goals in a classroom setting could also produce the desired results by emphasizing the educational importance of the study to the participants.

There are several limitations to this study that might have impacted its results. In order to ensure that the sample contained a baseline understanding of Among Us' rules and objectives, the recruitment materials and pre-screening questionnaire included text stating that players must have prior experience

playing the game to participate. While none of these materials explicitly stated which game titles would or would not be used during the study, several participants from both conditions expressed an expectation of playing Among Us prior to their session date. Such expectations could have invoked compensatory rivalry (Conrad & Conrad, 2005) or in some other way affected their engagement with their assigned intervention (e.g. DoomScroll players overlooking the modded elements, or tetris players feeling upset at being misled). In addition, multiple tetris sessions were conducted with smaller groups (i.e. 1-3 participants) in a smaller laboratory setting, whereas all DoomScroll sessions included 7-11 participants per session and were held in the UCI esports arena. While such deviations from the standard test environment were necessary in order to achieve the target sample size under a reasonable data collection timeline, these differences could have influenced players' mindsets when completing the interventions or assessments.

Chapter 6: Conclusion

In this dissertation, I have described the design, development, and evaluation of the Among Us mod DoomScroll as an intervention against misinformation. Games have proven to be a promising medium for inoculating individuals towards common tactics used for the spread of misinformation. However, few of the games currently in circulation for this purpose have been rigorously tested and even fewer have shown robust findings suggesting long-term efficacy. I attempted to resolve this issue by creating a game instilled with multiplayer gameplay and replayable narratives, two factors historically found to increase player retention. By choosing to mod a popular game title rather than develop our own, my team and I had hoped that we could expedite the development process by adopting an existing gameplay loop, assets, multiplayer infrastructure, and a dedicated audience. While we did reap these benefits and receive positive feedback and interest from players, modding ultimately produced logistic and creative barriers to our work. Our development team was small, made up of just a graduate student designer, a graduate student programmer, and an undergraduate programmer. Even with the latter two members possessing previous experience coding in C#, learning how to implement that code within the modding software was a complicated process in itself. Having any progress made then unexpectedly disrupted by updates from the primary developer led to frequent frustration and backtracking to find and fix any features that had been broken. I made several attempts to reach out to Innersloth to both alert them to our project and to ask for any assistance they might be willing to provide, but was unable to receive a response. These issues came to a head towards the end of playtesting, when several factors led to a pivotal decision that would go on to influence the overall direction of the project. By the ninth playtest, developer updates had produced delays that had caused us to run several months' past the scheduled deadline for beginning final data collection, to a point where my collaborators had other obligations they needed to turn their focus towards and where we knew that another update from Innersloth was imminent. Under this context and with the positive descriptive findings received from the final playtest group, I made the

decision to end development and move on to the randomized control trial. Further playtesting might have revealed that the outlier from that final test group actually represented the variance of efficacy inherent within social learning, and further iteration might have allowed us to better integrate the educational materials for a more uniform educational experience amongst players.

And so, while this project was successful in its own right along a number of metrics, I would make the following recommendations to educational game designers interested in attempting a similar path. Before beginning development, they must make serious considerations regarding the scale and feasibility of their design as well as the financial and temporal parameters under which they will operate. They must carefully consider which game title (and developing studio) they choose to mod and how its infrastructure might help or hinder their efforts. While modding can speed up the development cycle by eliminating the need for asset and system designs, the additional time needed for programmers to adopt the necessary modding software and adjust to unexpected changes by the developers can easily negate that benefit. Preserving this advantage might therefore require investing more resources into a larger programming team working under flexible deadlines in order to keep up with the pace of production, especially when working on a live online game with frequent patches. And while modding can allow for designers to focus their efforts more directly on the intended learning outcomes, they must also account for how the base game might inadvertently distract from the added features. Many players will likely have an inherent interest in the mod due to prior experience with its core title, but this experience will shape their approach and perspective with preconceived notions regarding how it should be played. This makes it vitally important for designers to deeply integrate their mechanics into the gameplay loop so that players are forced to engage with the new material in order to succeed. In these ways, modding can result in a longer and more complicated development process than creating a brand new game in which the team has full control. Evaluating these factors as early as possible will allow designers to determine whether they can realistically utilize modding to its fullest potential as a medium for serious games.

The randomized control trial revealed that the mod lacked quantifiable efficacy as an intervention, as DoomScroll players seemed to show no significant improvement in their discernment abilities when compared to tetris players. This could have been caused by several different factors. The mod was designed to be played casually for entertainment purposes, with its design goals hidden within the instructions and its educational materials integrated as complementary content rather than as the primary focus. This may have allowed players to avoid engaging with the educational aspects to an impactful degree. Similarly, the experimental protocol used might not have sufficiently directed the participants' focus towards the intervention's core content, or the assessments assigned might not have properly reflected the information that was actually being taught. It may also simply be that a collaboration-based gameplay loop is not conducive to direct transfer learning – "specifically predefined, concrete, and easily measurable learning objectives" (Harteveld & Bekebrede, 2011, pp. 52). Some players might have interacted with the educational content more than others and explained the lessons to a degree under which the group as a whole could succeed, but not to the extent that everyone would be able to apply the lessons on their own.

However, while DoomScroll might have struggled in uniformly teaching these concrete indicators of misinformation, it did appear to have success in other ways. Its design did result in positive reviews and notable player interest in future gameplay sessions, with much of the praise specifically referencing the added complexity and collaboration that was perceived to be missing from the base game. In this way, we were able to meet our design goals and create a misinformation intervention game that could conceivably draw players back in for repeated play after their first dose is administered. While further testing is necessary to confirm any longitudinal effects, I argue that DoomScroll possesses a realistic path towards long-term impacts once its educational materials are polished and more effectively integrated.

In the evaluated iteration, DoomScroll did also produce positive impacts in terms of open-ended learning (Harteveld & Bekebrede, 2011) by effectively teaching players to invoke the broader process of

civic online reasoning during gameplay. Researcher observations, post-match interviews, and in-game chat logs all suggested that players were finding it beneficial to actively investigate and validate the information provided by both the added evidence sources and their peers' own personal recollections. These acts of investigation mirror the techniques commonly used by professional fact-checkers towards real-life misinformation, conceivably preparing players to better assess potential falsehoods in their everyday lives. What's more, they were able to practice this process under novel contexts and realistic social conditions across multiple playthroughs, advantages that single-player, linear narrative games cannot provide. Thus, under these qualitative measures, DoomScroll did succeed as an intervention.

Based on these findings, I recommend that misinformation researchers, particularly those developing game-based inoculation interventions, strive to take greater advantage of the affordances that games as a medium can provide. Creating more realistic environments for individuals to hone their discernment abilities in can improve the likelihood of those skills transferring to the real world, better preparing them to resist misinformation when it occurs. Implementing avenues for meaningful interaction between players can both improve long-term player retention and better prepare them for navigating the biases and social pressures inherent in situations where their relationships with the sources are more complex than those found in fictional settings. Researchers must then also consider whether uniform quantitative assessments can sufficiently account for such social nuance when evaluating their game's impact. Misinformation can be spread in a variety of ways, under a variety of contexts, by sources with a variety of potential motivations, with the tactics used constantly being adapted to avoid detection. Teaching players not just what current techniques to watch out for but how to work with others to identify new techniques when they arise, could be the most effective way to collectively combat misinformation at the individual level on a global scale.

References

Agarwal, N., & Bandeli, K. K. (2017). Blogs, fake news, and information activities. *Digital Hydra: Security Implications of False Information Online*, 31-46.

Aguilar, E. M., & De La Vega, L. S. (2021, December). Missed Information: A Video Game Designed to Teach Methods of Spotting Fake News in Social Media. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)* (pp. 428-435). IEEE.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36.

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2), 2053168019848554.

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), eaay3539.

Al-Washmi, R., Bana, J., Knight, I., Benson, E., Kerr, O. A. A., Blanchfield, P., & Hopkins, G. (2014, October). Design of a math learning game using a Minecraft mod. In *European conference on games based learning* (Vol. 1, p. 10). Academic Conferences International Limited.

Arsenijević, J., & Andevski, M. (2016). New media literacy within the context of socio-demographic characteristics. *Procedia Technology*, 22, 1142-1151.

Arsenault, A., & Castells, M. (2006). Conquering the minds, conquering Iraq: The social production of misinformation in the United States–a case study. *Information, Communication & Society*, 9(3), 284-307.

Bad News Game (PC version) [Video game]. (2019) . DROG.

Baines, D., & Elliott, R. J. (2020). Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. *Discussion Papers*, 20(06), 20-06.

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication monographs*, 77(3), 281-311.

Barzilai, S., Mor-Hagani, S., Abed, F., Tal-Savir, D., Goldik, N. A., Talmon, I., & Davidow, O. (2023). Misinformation Is Contagious: Middle school students learn how to evaluate and share information responsibly through a digital game. *Computers & Education*, 202, 104832.

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of cognition*, 3(1).

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, 8(1), 20539517211013868.

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A signal detection approach to understanding the identification of fake news. *Perspectives on Psychological Science*, 17(1), 78-98.

Bauer, P. C., & Clemm von Hohenberg, B. (2021). Believing and sharing information by fake sources: an experiment. *Political communication*, 38(6), 647-671.

Benham, J. (2020). Best practices for journalistic balance: Gatekeeping, imbalance and the fake news era. *Journalism Practice*, 14(7), 791-811.

Berger, J., & Milkman, K. L. (2012). What makes online content viral?. *Journal of marketing research*, 49(2), 192-205.

Bergstrom, C., & West, J. (2017). Visualization: Misleading Axes on Graphs.

Berinsky, A. J. (2012). Rumors, truths, and reality: A study of political misinformation. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA.

Bless, H., Fiedler, K., & Forgas, J. P. (2006). Mood and the regulation of information processing and behavior. *Affect in social thinking and behavior*, 6584.

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4), 619-638.

Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, 65(5), 745-769.

Bouygues, H. L. (2018, December 4). A new front in the war against fake news: Kids. A New Front In The War Against Fake News: Kids. Retrieved May 2, 2022, from https://www.forbes.com/sites/helenleebouygues/2018/12/02/a-new-front-on-the-war-against-fake-news-kids/?sh=1bdd2bc8fce8

Boykoff, M. T., & Boykoff, J. M. (2007). Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum*, 38(6), 1190-1204.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318.

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5).

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation (Doctoral dissertation, University of Oxford).

Breuer, J., & Bente, G. (2010). Why so serious? On the relation of serious games and learning. *Journal for computer game culture*, 4, 7-24.

Brinol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European review of social psychology*, 20(1), 49-96.

Britt, M. A., Rouet, J. F., Blaum, D., & Millis, K. (2019). A reasoned approach to dealing with fake news. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 94-101.

Brumfiel, G. (2021, September 14). This doctor spread false information about COVID. she still kept her medical license. NPR. Retrieved May 2, 2022, from https://www.npr.org/sections/health-shots/2021/09/14/1035915598/doctors-covid-misinformation-medical-license

Brumfiel, G. (2021, September 14). This doctor spread false information about COVID. she still kept her medical license. NPR. Retrieved May 2, 2022, from https://www.npr.org/sections/health-shots/2021/09/14/1035915598/doctors-covid-misinformation-medical-license

Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, 87, 35-41.

Carnahan, D., Hao, Q., Jiang, X., & Lee, H. (2018). Feeling fine about being wrong: The influence of self-affirmation on the effectiveness of corrective information. *Human Communication Research*, 44(3), 274-298.

Carnahan, D., Bergan, D. E., & Lee, S. (2021). Do corrective effects last? Results from a longitudinal experiment on beliefs toward immigration in the US. *Political Behavior*, 43(3), 1227-1246.

Caulfield, M. (2017). Web literacy for student fact-checkers.

ChamberBreaker (PC version) [Video game]. (2021) . Jeon, Y., Kim, B., Xiong, A., Lee, D., & Han, K..

Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science*, 28(11), 1531-1546.

Chang, C. (2015). Motivated processing: How people perceive news covering novel or contradictory health research findings. *Science Communication*, 37(5), 602-634.

Chang, Y. K., Literat, I., Price, C., Eisman, J. I., Gardner, J., Chapman, A., & Truss, A. (2020). News literacy education in a polarized political climate: How games can teach youth to spot misinformation. *Harvard Kennedy School Misinformation Review*.

Chen, X. (2016). The influences of personality and motivation on the sharing of misinformation on social media. *IConference 2016 Proceedings*.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9).

Clarke, C. E. (2008). A question of balance: The autism-vaccine controversy in the British and American elite press. *Science communication*, 30(1), 77-107.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., ... & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 42(4), 1073-1095.

Clever, L., Assenmacher, D., Müller, K., Seiler, M. V., Riehle, D. M., Preuss, M., & Grimme, C. (2020, October). Fakeyou!-a gamified approach for building and evaluating resilience against fake news. In *Multidisciplinary International Symposium on Disinformation in Open Online Media* (pp. 218-232). Springer, Cham.

CO2peration (PC Version) [Board game]. (2020).

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of personality and social psychology*, 85(5), 808.

Compton, J. A., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, 29(1), 97-146.

Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post‐truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass*, 15(6), e12602.

Conrad, K. M., & Conrad, K. J. (2005). Compensatory rivalry. *Encyclopedia of statistics in behavioral science*.

Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PloS one*, 12(5), e0175799.

Cranky Uncle (PC version) [Video game]. (2020) . Monash University.

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature human behaviour*, 1(11), 769-771.

Dame Adjin-Tettey, T. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent arts & humanities*, 9(1), 2037229.

DellaVigna, S., & Kaplan, E. (2007). The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3), 1187-1234.

DeJong, S. (2023). Playing With Fake News: state of fake news video games.

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, 1(1).

Di Domenico, G., Nunan, D., Sit, J., & Pitardi, V. (2021a). Free but fake speech: When giving primacy to the source decreases misinformation sharing on social media. *Psychology & Marketing,* 38(10), 1700-1711.

Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021b). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329-341.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of learning and motivation*, 50, 307-338.

Ecker, U. K., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short‐format refutational fact‐checks. *British Journal of Psychology*, 111(1), 36-54.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13-29.

Eichenberg, C., & Schott, M. (2017). Serious games for psychotherapy: A systematic review. *Games for Health Journal*, 6(3), 127-135.

Einwiller, S. A., & Kim, S. (2020). How online content providers moderate user‐generated content to prevent harmful online communication: An analysis of policies and their implementation. *Policy & Internet*, 12(2), 184-206.

Ellis, H. A. & Ashbrook, P. W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fielder & J. Forgas (ed.), *Affect, cognition, and social behavior* (pp. 25-43) . Hogrefe .

Escape the Fake (PC version) [Video game]. (2021) . Paraschivoiu, I., Buchner, J., Praxmarer, R., & Layer-Wagner, T..

Factitious (PC version) [Video game]. (2018) . American University.

Fake it to Make it (PC version) [Video game]. (2018) . Amanda Warner.

Fakeyou! (PC version) [Mobile game]. (2021) . Clever, L., Assenmacher, D., Müller, K., Seiler, M. V., Riehle, D. M., Preuss, M., & Grimme, C..

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993.

Fazio, M. (2021, April 27). How mask guidelines have evolved. The New York Times. Retrieved May 2, 2022, from https://www.nytimes.com/2021/04/27/science/face-mask-guidelines-timeline.html

Fenker, D. B., Waldmann, M. R., & Holyoak, K. J. (2005). Accessing causal relations in semantic memory. *Memory & cognition*, 33(6), 1036-1046.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9(2), 319–342. https://doi.org/10.1177/1461444807075015

Fleming, T. M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S. N., Cheek, C., ... & Hetrick, S. (2017). Serious games and gamification for mental health: current status and promising directions. *Frontiers in psychiatry*, 7, 215.

Forgas, J. P. (2019). Happy believers and sad skeptics? Affective influences on gullibility. *Current Directions in Psychological Science*, 28(3), 306-313.

Frattesi, T., Griesbach, D., Leith, J., Shaffer, T., & DeWinter, J. (2011). Replayability of video games. *IQP*, Worcester Polytechnic Institute, Worcester.

Fuster, H., Chamarro, A., Carbonell, X., & Vallerand, R. J. (2014). Relationship between passion and motivation for gaming in players of massively multiplayer online role-playing games. *Cyberpsychology*, Behavior, and Social Networking, 17(5), 292-297.

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016, June). Social clicks: What and who gets read on Twitter?. In *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science* (pp. 179-192).

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & gaming*, 33(4), 441-467.

Gee, J. P. (2004). Learning by design: Games as learning machines. *Interactive educational multimedia: IEM*, 15-23.

Gervais, W. M. (2015). Override the controversy: Analytic thinking predicts endorsement of evolution. *Cognition*, 142, 312-321.

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, 336(6080), 493-496.

GoViral! (PC version) [Video game]. (2021) . DROG.

Grace, L., & Liang, S. (2023). Examining Misinformation and Disinformation Games Through Inoculation Theory and Transportation Theory.

Grady, R. H., Ditto, P. H., & Loftus, E. F. (2021). Nevertheless, partisanship persisted: fake news warnings help briefly, but bias returns with time. *Cognitive research: principles and implications*, 6(1), 1-16

Graham, S. (2014). Rolling your own: On modding commercial games for educational goals. *Pastplay: Teaching and learning history with technology*, 214-227.

Grasso, M. (2020). Towards a broader climate ethics: Confronting the oil industry with morally relevant facts. *Energy Research & Social Science*, 62, 101383.

Groeling, T. (2008). Who's the fairest of them all? An empirical test for partisan bias on ABC, CBS, NBC, and Fox News. *Presidential Studies Quarterly*, 38(4), 631-657.

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances*, 5(1), eaau4586.

Guess, A., & Coppock, A. (2020). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, 50(4), 1497-1515.

Hameleers, M. (2022). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, 25(1), 110-126.

Harker-Schuch, I. E., Mills, F. P., Lade, S. J., & Colvin, R. M. (2020). CO2peration–Structuring a 3D interactive digital game to improve climate literacy in the 12-13-year-old age group. *Computers & Education*, 144, 103705.

Harmony Square (PC version) [Video game]. (2020) . DROG.

Harteveld, C., & Bekebrede, G. (2011). Learning in single-versus multiplayer games: The more the merrier?. *Simulation & Gaming*, 42(1), 43-63.

Hoadley, C. P. (2002). Creating context: Design-based research in creating and understanding CSCL.

Hobbs, R. (2010). Digital and Media Literacy: A Plan of Action. A White Paper on the Digital and Media Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a Democracy. *Aspen Institute*. 1 Dupont Circle NW Suite 700, Washington, DC 20036.

Horne, B., & Adali, S. (2017, May). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media* (Vol. 11, No. 1, pp. 759-766).

Hughes, M. G., Griffith, J. A., Zeni, T. A., Arsenault, M. L., Cooper, O. D., Johnson, G., ... & Mumford, M. D. (2014). Discrediting in a message board forum: The effects of social support and attacks on expertise and trustworthiness. *Journal of Computer-Mediated Communication*, 19(3), 325-341.

Hunt, E. (2017, January 22). Trump's inauguration crowd: Sean Spicer's claims versus the evidence. The Guardian. Retrieved April 27, 2022, from https://www.theguardian.com/us-news/2017/jan/22/trump-inauguration-crowd-sean-spicers-claims-versus-the-evidence.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research*, 49(10), 1295-1306.

Ivanov, B., Parker, K. A., & Dillingham, L. L. (2018). Testing the limits of inoculation-generated resistance. *Western Journal of Communication*, 82(5), 648-665.

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences*, 116(16), 7656-7661.

Jacobson, G. C. (2010). Perception, memory, and partisan polarization on the Iraq War. *Political science quarterly*, 125(1), 31-56.

Jeon, Y., Kim, B., Xiong, A., Lee, D., & Han, K. (2021). ChamberBreaker: mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-26.

Jia, F. (2020). Misinformation Literature Review: Definitions, Taxonomy, and Models. *International Journal of Social Science and Education Research*, 3(12), 85-90.

Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PloS one*, 9(2), e89177.

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti‐vaccine conspiracy theories. *Journal of Applied Social Psychology*, 47(8), 459-469.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, 2(10), 732-735.

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition.

Katsaounidou, A., Vrysis, L., Kotsakis, R., Dimoulas, C., & Veglis, A. (2019). MAthE the game: A serious game for education and training in news verification. *Education Sciences*, 9(2), 155.

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, 51(5-6), 374-397.

Klopfer, E., & Squire, K. (2008). Environmental Detectives—the development of an augmented reality platform for environmental simulations. *Educational technology research and development*, 56, 203-228.

Kou, Y., & Gui, X. (2014, October). Playing with strangers: understanding temporary teams in league of legends. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (pp. 161-169).

Kücklich, J., & Fellow, M. C. (2004). Play and playability as key concepts in new media studies. *STeM Centre*, Dublin City University.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480.

Lamboozled! (PC version) [Video game]. (2019) . MASCLab.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Lepper, M. R., & Malone, T. W. (2021). Intrinsic motivation and instructional effectiveness in computer-based education. In *Aptitude, learning, and instruction* (pp. 255-286). Routledge.

Levendusky, M. (2013). Partisan media exposure and attitudes toward the opposition. *Political communication*, 30(4), 565-581.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of experimental social psychology*, 46(6), 1093-1096.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3), 106-131.

Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. *Current Directions in Psychological Science*, 25(4), 217-222.

Lewandowsky, S., Cook, J., Ecker, U., Albarracin, D., Amazeen, M., Kendou, P., ... & Zaragoza, M. (2020). The debunking handbook 2020.

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 1-38.

Li, M. H., Chen, Z., & Rao, L. L. (2022). Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior*, 107295.

Lima, C. (2020, August 31). Twitter changing labeling practices after deceptive videos hit Biden. POLITICO. Retrieved April 27, 2022, from https://www.politico.com/news/2020/08/31/twitter-doctored-video-policy-biden-406632

Literat, I., Chang, Y. K., Eisman, J., & Gardner, J. (2021). LAMBOOZLED!: The design and development of a game-based approach to news literacy education. *Journal of Media Literacy Education*, 13(1), 56-66.

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, 5(1), 86-88.

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature human behaviour*, 5(3), 337-348.

Loos, E., & Nijenhuis, J. (2020). Consuming Fake News: A Matter of Age? The perception of political fake news stories in Facebook ads. In Human Aspects of IT for the Aged Population. *Technology and Society: 6th International Conference, ITAP 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part III 22* (pp. 69-88). Springer International Publishing.

Lu, A. S., Kharrazi, H., Gharghabi, F., & Thompson, D. (2013). A systematic review of health videogames on childhood obesity prevention and intervention. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, 2(3), 131-141.

Lubens, P. (2015). Journalists and public health professionals: Challenges of a symbiotic relationship. *Disaster medicine and public health preparedness*, 9(1), 59-63.

Lyons, T. (2017, December 20). Replacing disputed flags with related articles. Replacing Disputed Flags With Related Articles. Retrieved April 27, 2022, from https://about.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1.

Maertens, R., Götz, F. M., Golino, H. F., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., ... & van der Linden, S. (2024). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods*, 56(3), 1863-1899.

MathE (PC version) [Video game]. (2019) . Katsaounidou, A., Vrysis, L., Kotsakis, R., Dimoulas, C., & Veglis, A. .

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, 5(1), 1-20.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological science*, 11(5), 424-428.

McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, 24(2), 184-197.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165-193.

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: An intervention study. *British Journal of Educational Psychology*, 89(3), 485-500.

McGrew, S., & Breakstone, J. (2023). Civic Online Reasoning Across the Curriculum: Developing and Testing the Efficacy of Digital Literacy Lessons. *AERA Open*, 9, 23328584231176451.

Micallef, N., Avram, M., Menczer, F., & Patil, S. (2021). Fakey: A Game Intervention to Improve News Literacy on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-27.

Missed Information (PC version) [Video game]. (2021) . Aguilar, E. M., & De La Vega, L. S..

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" is not simply false information: A concept explication and taxonomy of online content. *American behavioral scientist*, 65(2), 180-212.

Mosleh, M., Pennycook, G., & Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *Plos one*, 15(2), e0228882.

Mustaro, P. N., & Mendonça, R. L. (2012). Immersion, narrative, and replayability as the motivational and attractiveness factors in serious games. *In Handbook of research on serious games as educational, business and research tools* (pp. 991-1008). IGI Global.

Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and simulated social media postings: an experimental investigation of source credibility, repeated exposure, and presentation format. *Cognitive Research: Principles and Implications*, 5(1), 1-16.

Nardi, B., & Harris, J. (2006, November). Strangers and friends: Collaborative play in World of Warcraft. *In Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 149-158).

Newsfeed Defender (PC version) [Video game]. (2018) . iCivics.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.

O'Connor, C., & Weatherall, J. O. (2019). The misinformation age. In The Misinformation Age. Yale University Press.

Pallavicini, F., Ferrari, A., & Mantovani, F. (2018). Video games for well-being: A systematic review on the application of computer games for cognitive and emotional training in the adult population. *Frontiers in psychology*, 9, 2127.

Pakarinen, A., Parisod, H., Smed, J., & Salanterae, S. (2017). Health game interventions to enhance physical activity self‐efficacy of children: a quantitative systematic review. *Journal of advanced nursing*, 73(4), 794-811.

Paraschivoiu, I., Buchner, J., Praxmarer, R., & Layer-Wagner, T. (2021, October). Escape the Fake: Development and Evaluation of an Augmented Reality Escape Room Game for Fighting Fake News. *In Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play* (pp. 320-325).

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development*, 60, 723-751.

Pavleska, T., Školkay, A., Zankova, B., Ribeiro, N., & Bechmann, A. (2018). Performance analysis of fact-checking organizations and initiatives in Europe: a critical overview of online platforms fighting fake news. *Social media and convergence*, 29, 1-28.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123(3), 335-346.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425-432.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, 147(12), 1865.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39-50.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality*, 88(2), 185-200.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in cognitive sciences*.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, 31(7), 770-780.

Persson, E., Andersson, D., Koppel, L., Västfjäll, D., & Tinghög, G. (2021). A preregistered replication of motivated numeracy. *Cognition*, 214, 104768.

Pfau, M., Bockern, S. V., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communications Monographs*, 59(3), 213-230.

Play Fake News (PC version) [Video game]. (2017) . ISL.

Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: failure of three pro-vaccination strategies. *PloS one*, 12(7), e0181640.

Pilegard, C., & Mayer, R. E. (2016). Improving academic learning from computer-based narrative games. *Contemporary Educational Psychology*, 44, 12-20.

Pine, R., Fleming, T., McCallum, S., & Sutcliffe, K. (2020). The effects of casual videogames on anxiety, depression, stress, and low mood: a systematic review. *Games for health journal*, 9(4), 255-264.

Polititruth (PC version) [Video game]. (2018) . Politifact.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology*, 34(2), 243-281.

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26).

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and cognition*, 8(3), 338-342.

Readfearn, G. (2016). Revealed: Most popular climate story on social media told half a million people the science was a hoax. *Desmog*, 29.

Roozenbeek, J., & Van der Linden, S. (2019a). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 1-10.

Roozenbeek, J., & Van Der Linden, S. (2019b). The fake news game: actively inoculating against the risk of misinformation. *Journal of risk research*, 22(5), 570-580.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., ... & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, 7(10), 201199.

Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. *The Harvard Kennedy School Misinformation Review*.

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., Sultan, M., & van der Linden, S. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making*, 17(3), 547-573.

Sanderson, Z., Brown, M. A., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Twitter flagged Donald Trump's tweets with election misinformation: They continued to spread both on and off the platform. *Harvard Kennedy School Misinformation Review*.

Schaffner, B. F., & Luks, S. (2018). Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys. *Public Opinion Quarterly*, 82(1), 135-147.

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., ... & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology*.

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, 2(1), 85-95.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96, 104.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3), 1-42.

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New media & society*, 21(2), 438-463.

Shen, C., Wang, H., & Ritterfeld, U. (2009). Serious games and seriously fun games: Can they be one and the same?. *In Serious Games* (pp. 70-84). Routledge.

Shi, W. (2023). Efficacy of Educational Misinformation Games. arXiv preprint arXiv:2305.09429.

Silverman, C., Timberg, C., Kao, J., & Merrill, J. B. (2022, January 4). Facebook groups topped 10,000 daily attacks on election before Jan. 6, analysis shows. The Washington Post. Retrieved May 3, 2022, from https://www.washingtonpost.com/technology/2022/01/04/facebook-election-misinformation-capitol-riot/

Sirlin, N., Epstein, Z., Arechar, A. A., & Rand, D. (2021). Digital literacy and susceptibility to misinformation.

Skurnik, I., Yoon, C., & Schwarz, N. (2007). Education about flu can reduce intentions to get a vaccination. Unpublished manuscript.

Smith, P., Bansal-Travers, M., O'Connor, R., Brown, A., Banthin, C., Guardino-Colket, S., & Cummings, K. M. (2011). Correcting over 50 years of tobacco industry misinformation. *American journal of preventive medicine*, 40(6), 690-698.

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social cognition*, 26(6), 791-799.

Squire, K. (2003). Video games in education. Int. J. Intell. *Games & Simulation*, 2(1), 49-62.

Steinkuehler, C. A. (2004). Learning in massively multiplayer online games.

Steinkuehler, C., Anderson, C. G., Reitman, J. G., Lee, J. S., Wu, M., Wells, G., & Gardner, R. T. (2023). Enriched Esports: The Design and Four-Year Examination of a School-Affiliated Competitive Videogame Program for Youth. *Journal of Interactive Learning Research*, 34(1), 59-119.

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., ... & Loker, K. (2019). Who shared it?: Deciding what news to trust on social media. *Digital journalism*, 7(6), 783-801.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview.

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133(3), 572-585.

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020a). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, 9(3), 286

Swire‐Thompson, B., Ecker, U. K., Lewandowsky, S., & Berinsky, A. J. (2020b). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, 41(1), 21-34.

Swire-Thompson, B., Cook, J., Butler, L. H., Sanderson, J. A., Lewandowsky, S., & Ecker, U. K. (2021). Correction format has a limited role when debunking misinformation. *Cognitive research: principles and implications*, 6(1), 1-15.

Swire-Thompson, B., Miklaucic, N., Wihbey, J. P., Lazer, D., & DeGutis, J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. *Journal of Experimental Psychology: General*.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, 31(2), 137-155.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3), 755-769.

Tan, A. S., & Bigman, C. A. (2020). Misinformation about commercial tobacco products on social media—Implications and research opportunities for reducing tobacco-related health disparities. *American journal of public health*, 110(S3), S281-S283.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism*, 6(2), 137-153.

Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. (2021). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology*.

The Fake News Game (PC version) [Video game]. (2019) . *DROG*.

Tormala, Z. L., & Petty, R. E. (2004). Source credibility and attitude certainty: A metacognitive analysis of resistance to persuasion. *Journal of Consumer Psychology*, 14(4), 427-442.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PloS one*, 13(9), e0203958.

Torres, R., Gerhart, N., & Negahban, A. (2018, January). Combating fake news: An investigation of information verification behaviors on social networking sites. *In Proceedings of the 51st Hawaii international conference on system sciences*.

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, 185, 111269.

Trustme! (PC version) [Video game]. (2018) . Yang, S., Lee, J. W., Kim, H. J., Kang, M., Chong, E., & Kim, E. M..

Tully, M., Vraga, E. K., & Bode, L. (2020). Designing and testing news literacy messages for social media. *Mass Communication and Society*, 23(1), 22-46.

Urban, A., Hewitt, C., & Moore, J. (2018). Fake it to make it, media literacy, and persuasive design: Using the functional triad as a tool for investigating persuasive elements in a fake news simulator. *Proceedings of the Association for Information Science and Technology*, 55(1), 915-916.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in cognitive sciences*, 22(3), 213-224.

Van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: plain facts, pie charts or metaphors?. *Climatic Change*, 126(1), 255-262.

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008.

Van der Linden, S., Leiserowitz, A., & Maibach, E. (2018). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour,* 2(1), 2-3.

Van der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 1-8.

Van der Linden, S. (2023). We need a gold standard for randomised control trials studying misinformation and vaccine hesitancy on social media. *bmj*, 381.

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

Vraga, E. K., & Bode, L. (2018). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21(10), 1337-1353.

Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication*, 37(1), 136-144.

Vraga, E. K., Kim, S. C., Cook, J., & Bode, L. (2020). Testing the effectiveness of correction placement and type on Instagram. *The International Journal of Press/Politics*, 25(4), 632-652.

Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., ... & Walker-Smith, J. A. (1998). RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children.

Wang, P., Angarita, R., & Renna, I. (2018, April). Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. In *Companion Proceedings of the The Web Conference 2018* (pp. 1557-1561).

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423-441.

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?. *Communication research*, 47(2), 155-177.

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking (Vol. 27, pp. 1-107). *Strasbourg: Council of Europe*.

Wardle, C. (2020). Understanding Information Disorder: Essential Guides. First Draft.

Wells, C., Reedy, J., Gastil, J., & Lee, C. (2009). Information distortion and voting choices: The origins and effects of factual beliefs in initiative elections. *Political Psychology*, 30(6), 953-969.

Wells, G. (2023, November 15). Doom Scroll: Modding Among Us to Improve Misinformation Discernment. Retrieved from osf.io/9hqcu

Wendel, V., Hertin, F., Göbel, S., & Steinmetz, R. (2010). Collaborative learning by means of multiplayer serious games. In *Advances in Web-Based Learning–ICWL 2010: 9th International Conference*, Shanghai, China, December 8-10, 2010. Proceedings 9 (pp. 289-298). Springer Berlin Heidelberg.

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. Proceedings of the *National Academy of Sciences*, 118(15), e1912444117.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11).

Wineburg, S., & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information.

Winkielman, P., Huber, D. E., Kavanagh, L., & Schwarz, N. (2012). Fluency of consistency: When thoughts fit nicely and flow smoothly. *Cognitive consistency: A fundamental principle in social cognition*, 89-111.

World Economic Forum. (2024, January 10). Global risks 2024: Disinformation tops global risks 2024 as environmental threats intensify. weforum.org. https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, 21(2), 80-90.

Yang, S., Lee, J. W., Kim, H. J., Kang, M., Chong, E., & Kim, E. M. (2021). Can an online educational game contribute to developing information literate citizens?. *Computers & Education*, 161, 104057.

Yousuf, H., van der Linden, S., Bredius, L., van Essen, G. T., Sweep, G., Preminger, Z., ... & Hofstra, L. (2021). A media intervention applying debunking versus non-debunking content to combat vaccine misinformation in elderly in the Netherlands: A digital randomised trial. *EClinicalMedicine*, 35, 100881.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... & Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual differences*, 48(8), 889-894.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality* (JDIQ), 11(3), 1-37.

Appendix A. Demographic Questions for Playtest Survey

Demographic Factors (Pre-test only)

1. Please state your age (in years): [numeric response]
2. Please state your gender: [Drop down: Male, Female, Nonbinary, Prefer not to answer, Prefer to self-describe: ( )]
3. Please describe your education level: [Drop down: High School, Some College, Bachelor's Degree, Postgraduate Degree]
4. On a scale from 1 to 5 (1 meaning Very Liberal, 3 being Neutral, and 5 meaning Very Conservative), please describe your political leaning: [5-point likert]
5. Have you ever played Among us before? [Yes/No]
    a. [If yes] How many hours would you estimate you have played? [numeric response]
6. Have you played modded versions of Among Us before? [Yes/No]
    a. [If yes] How many hours would you estimate you have played? [numeric response]

User-test Questionnaire (Post-test only)

1. How many matches did you complete during this play session prior to this survey?
    a. 1
    b. 2
    c. 3
    d. 4
    e. 5+
2. On average, how many other players were included in your matches during this play session?
    a. 4 through 15 players
3. On a scale of 1 to 5 (1 being "never", 3 being neutral, and 5 being "often"), rate your usage of the DoomScroll mod mechanics during your play session. These mechanics include the static chat logs, the secondary win conditions, the task sign-in function, the screenshot function, the headlines feature, and the folders during meetings. [likert]
4. On a scale of 1 to 5 (1 being "Not confident at all"", 3 being "Somewhat confident", and 5 being "Very confident"), rate your confidence in your ability to identify misinformation. [likert]
5. Do you feel more confident in your ability to identify misinformation after playing? [Yes/No]
    a. If no, why not? [free response]
6. On a scale of 1 to 5 (1 being "Strongly disagree", 3 being Neutral, and 5 being "Strongly Agree"), rate the following statements: [likert]
    a. I enjoyed playing Among Us using the mod.
    b. I would play Among Us using the mod again.
    c. I found the **static chat log** mechanic **fun**.
    d. I found the **secondary win conditions** mechanic **fun**.
    e. I found the **task sign-in** mechanic **fun**.
    f. I found the **folder menu** mechanic **fun**.
    g. I found the **screenshot** mechanic **fun**.
    h. I found the **headline** mechanic **fun**.

    i. I found the **static chat log** mechanic **useful**.
    j. I found the **secondary win conditions** mechanic **useful**.
    k. I found the **task sign-in** mechanic **useful**.
    l. I found the **folder menu** mechanic **useful**.
    m. I found the **screenshot** mechanic **useful**.
    n. I found the **headline** mechanic **useful**.
7. In your own words, how would you describe your experience using the modded mechanics (secondary win conditions, task sign-ins, folder menu during meetings)? [free response]
8. In your opinion, did the modded mechanics change how you normally play Among Us? If yes, please explain. [free response]
9. During your play session, did you experience any issues with the modded mechanics? [free response]
10. In your opinion, how might the modded features be improved? [free response]

Appendix B. Misinformation Susceptibility Test (MIST-8)

Please categorize the following news headlines as either 'Fake News' or 'Real News'. (Both Pre- and Post-)
Some items may look credible or obviously false at first sight, but may actually fall in the opposite category. However, for each news headline, only one category is correct.

1. Government Officials Have Manipulated Stock Prices to Hide Scandals
2. New Study: Left-Wingers Are More Likely to Lie to Get a Higher Salary
3. Certain Vaccines Are Loaded with Dangerous Chemicals and Toxins
4. The Government Is Knowingly Spreading Disease Through the Airwaves and Food Supply
5. Attitudes Toward EU Are Largely Positive, Both Within Europe and Outside It
6. Hyatt Will Remove Small Bottles from Hotel Bathrooms by 2021
7. Republicans Divided in Views of Trump’s Conduct, Democrats Are Broadly Critical
8. Global Warming Age Gap: Younger Americans Most Worried

Appendix C. Civic Online Reasoning Assessments

***News on Twitter* Assessment.**

The following tweets appeared on March 11, 2015.

What's happening?

The Subject @handsupunited_ • 5h
#Ferguson Police Chief Resigns! #ShutItDown

46 39 View photo

Lisa Bloom @LisaBloom • 5h
#Ferguson police chief steps down, to be replaced by an interim chief from the same police department DOJ found is riddled with racial bias.

150 71

NPR @NPR • 6h
#Ferguson police chief's resignation is effective March 19 via @MilesParks @nprnews: n.pr/1E6zF4j

236 115

Bassem Masri @bassem_masri • 7h
Supposedly Chief Jackson is resigning today but we will have more info later **#Ferguson**

Which tweet is the best source of information about the police chief's resignation? Explain why.

Claims on Social Media **Assessment.**

The following tweet appears in your Twitter feed:
https://twitter.com/MoveOn/status/666772893846675456?lang=en

Why might this tweet be a useful source about NRA members' opinions on background checks? List any sources you used to make your decision.

Researching a Claim **Assessment.**

Some people claim that Margaret Sanger, the founder of Planned Parenthood, supported euthanasia. Take about 8 minutes doing research online to decide if you believe this claim is true.

Do you believe Margaret Sanger supported euthanasia? Explain using evidence from the websites you consulted.

Appendix D. Perceived Reliability Assessment

HEARTBREAKING: Baby penguin drowns in enclosure at zoo, fellow penguins cry for DAYS!

Animals News @AnimalsNews

Brazilian Official Caught Embezzling Millions of Dollars from Campaign Funds.

CNN @_CNN_

HUGE NEWS: Secret Group of NFT Millionaires Plot to Take Over Puerto Rico.

Daily News 365 @DailyNews365

Today, President Biden rejoined the Paris Agreement on Climate Change.

The White House @WhiteHouse

Seeing recent reports suggesting that a local hospital is serving as a front for a drug distribution ring.

Ken James @KJamesNews

Due to scheduling conflicts, Liam Hemsworth has withdrawn from the role of Geralt of Rivia in #TheWitcher season 4. We wish Liam nothing but the best.

Nétflix @nétflixstudios

As a scientist, I do not believe there is enough evidence to suggest that humans can influence the climate.

Steve Smith @SteveSEconPhD

Hey @LeoDiCaprio, it's snowing in New York. Could use some of that global warming you're always crying about!

Southpaw Goalkeeper @southpawgk

New study shows left-wing people lie more often than right-wing people.

Frontier News @FrontierNews

Study shows right-leaning individuals are less likely to donate to charity than left-leaning individuals

Mountain Media @MountainMedia

The Lakers are building a bronze statue of Kobe Bryant outside of Crypto Arena on March 6th.

Adrian Wojnarowski @wojespn

New research shows that pet ownership slows cognitive decline among older adults.

News Bot @newsbot

New research shows that pet ownership slows cognitive decline among older adults.

News Bot @newsbot

“Jawn”, “decision fatigue”, and “nepo baby” are among the over 500 new words and phrases added to the latest version of Dictionary.com, the site announced Wednesday.

The Associated Press @AP

Marvel Studios President Kevin Feige is reportedly “angry and embarrassed” at Disney over its legal battle with Scarlett Johansson.

IGN www.ign.com

Just watched an employee toss my bag across the tarmac, but at least there’s a plane this time! @SouthwestAir hates its customers

Eshaaaa @21Esha21

Children’s hospital employees caught ignoring cries of newborn infants, smoking in ICU. Parents horrified.

Lubbock Local News @LubbockNews

Drinking tea 3 times a day can help improve your heart health.

u/KyleM38 on reddit r/askUK

Over 2000 people have already died from this new strain of pneumonia.

Medical News Updates www.medicalnewsupdates.wordpress.com

BREAKING NEWS: Millions have fallen deathly ill after receiving this year's flu shot, are pro-vaccine doctors to blame?

Jessica Paxton @JPReports

Los Angeles could be completely underwater within the next 6 months due to global warming, experts say.

Progressive Network @ProgressiveNetwork

Appendix E. Demographic Questions for Randomized Control Trial.

Please state your age (in years): [numeric response]

Please state your gender: [Drop down: Male, Female, Nonbinary, Prefer not to answer, Prefer to self-describe: ( )]

Please describe your education level: [Drop down: High School, Some College, Bachelor's Degree, Postgraduate Degree]

On a scale from 1 to 5 (1 meaning Very Liberal, 3 being Neutral, and 5 meaning Very Conservative), please describe your political leaning: [5-point likert]

Have you ever played Among us before? [Yes/No]

[If yes] How many hours would you estimate you have played? [numeric response]

Have you played modded versions of Among Us before? [Yes/No]

During your play session, did you have any pre-existing relationships (friends, family, coworkers, etc.) with any of the other participants? [Yes/No]

Appendix F. A Priori Codes

A priori codes used for thematic analysis of qualitative RCT data. Improvement over time (e.g. increased accuracy scores or more frequent mention of common unreliability indicators, increased usage of features, increased reference to evidence sources while discussing arguments during meetings) indicates positive behavioral outcomes.

In-game indicators	**Learning Goals**	**Data Collection**	**References:**
Headlines: Accuracy through headline win condition (correctly discerning true vs. false headlines)	Common indicators of false/misleading information	in-game data (75% of crewmates must reach 100% accuracy to win match)	● Zannettou et al. (2019) ● Molina et al, 2021 ● Hameleers, 2022
Headlines: Supporting arguments/refuting opposing claims using indicators (e.g. in chat, pointing out misspelling as indicator of unreliable source)	Common indicators of false/misleading information	Observations, chat logs, in-game telemetry data	● Zannettou et al. (2019) ● Molina et al, 2021 ● Hameleers, 2022
Usage and recognition of common indicators/techniques of misinformation during meetings (e.g. discrediting or trolling another player, starting conspiracy theories, or making emotional or hyperbolic arguments)	Common indicators of false/misleading information	Observations, chat logs	● Zannettou et al. (2019) ● Molina et al, 2021 ● Wardle, 2020 ● Lischka, 2019
Investigation of evidence (posting tasks and headlines during gameplay, opening folders during meetings)	Lateral reading – investigating evidence sources	Observations, in-game telemetry data	● Wineberg & McGrew, 2017
Formulating arguments around evidence (sharing headlines and tasks to chat, discussing evidence sources in chat)	Lateral reading – utilizing evidence in discernment	Observations, chat logs, sharing headlines/tasks in chat	● Wineberg & McGrew, 2017

Appendix G. Headlines & Sources

Untrustworthy headlines (player-generated):

- "My cat trusts {X}, I trust {X}. Ms. Meatball can sense good auras"
- "{X} just saved my puppy from a burning building!"
- "{X} just saved the ship from sabotage #hero"
- "{X} is a superhero sent to protect us we luv {X} <3 <3 <3"
- "I can't believe {X} and {Y} are DATING #scandalous"
- "{X} is smart, funny, brilliant, can definitely kickflip"
- "{X} is complete mainstream propaganda fake news"
- "I can't believe anyone trusts {X}"
- "Anyone who believes what {X} says is a sheep, a baby baby lamb"
- "{X} seen stealing food from puppies!"
- "Is {X} really who they say they are?"
- "{X} Isn't doing their fair share!"
- "{X} just stole my lunch from the fridge :("
- "{X} KEPT ME UP ALL NIGHT! STOP PARTYING PLEASE"
- "{X} says {Y} hasn't done ANY tasks???? INSANE"
- "I just caught {X} loitering near vents???"
- "I've never seen {X} kickflip even ONCE"

Trustworthy headlines (news source-generated):

- "{X} Has Completed Several Tasks."
- "{X} has completed {#} tasks"
- "{X} has done {#} tasks so far this match."
- "{#} task sign-ins recorded for {X}."
- "{X} has previously signed in to {T}."
- "{X} has signed in to {#} tasks."
- "{X} Previously Voted for {Y}"
- "{X} is not currently a suspect", local investigators state."
- "{X} has prevented a sabotage attempt."
- "{X} Saves Ship from Sabotage."
- "Sabotage logs show {X} helped save the ship."
- "Security Records Reveal {X} Has Prevented {#S} Sabotages"
- "Records show {X} has completed 0 tasks thus far."
- "{X} has done {#} tasks so far this match."
- "Task Records Show {X} Hasn't Completed Their Tasks"
- "{X} is Missing from Sign-In Sheet"
- "{#} task sign-ins recorded for {X}."
- "{X} Seen Loitering Near Vents."
- "'{X} CANNOT Be Trusted', Local Ghost Warns."
- "Report: {X} Committed Sabotage."

Untrustworthy headlines (news source-generated):

- "HERO: {X} Saves Ship from Sabotage.",
- "{X} saves Puppy from Burning Building!",
- "Streets safe again: Our very own innocent {X} didn't kill anyone last round!!!"
- "{X} Does Tasks Quickly, Votes Correctly Every Time."
- "{X} Finishes Tasks at \"World Record Pace\", Says Self-Proclaimed Task Expert."
- "Art Professor: "As a doctor, I'm 100% certain that {X} is NOT an impostor.""
- "{X} Seen Stealing Food from Puppies!"
- "Is {X} Really Who They Say They Are?"
- "{X} Isn't Doing Their Fair Share!"
- "{X} Caught Stealing Lunches from Fridge."
- "{X} CANNOT Be Trusted', Local Ghost Warns."
- "{X} Parties 'Til 4am, Keeps Shipmates Awake."
- "LAZY??? Task Records Reveal {X} Hasn't Completed Their Tasks"
- "Imposter Among Us? {X} is Missing from Sign-In Sheet"
- "LIGHTS OUT: Security Records Suggest {X} Sabotaging Crew"
- "BUSY BEE: {X} Has Completed More Tasks Than {Y}."
- "{X} Caught Loitering Near Vents, "Like a Rat" Says Witness"
- "{X} and {Y} Caught Sharing Headphones at Lunch"
- "{X} seen arguing with {Y}, Calls Them "Smallest Bean""
- "{X} and {Y} Spotted Holding Hands During Meeting?!"
- "Anonymous Source: \"{X} is the imposter",
- "BREAKING: {X} has only signed in to {#} tasks all game"
- "{X} Seen Loitering Near Vents.",
- "{X} Frustrated, Blames {Y} for Slow Task Completion"
- "TENSION ON THE SHIP! {X} seen arguing with {Y}."
- "{X} Has Done Zero Tasks AND Has Not Voted, Suspicious?"
- "Economist: \"I diagnose {X} with the psychological profile of an impostor"

Untrustworthy sources (players):

- "{Color} Crewm8_",
- "@NotMe29",
- "@FreddyGazebo1922830485",
- "@xXCrewhateXx",
- "{Color} Grewmate",
- "@{Name}_1515386",
- "@REAL{Name}",
- "@{Name}official",
- "@{NameMispelling}",
- "@VersacePelligrino",
- "@JMinietian | Journalist @SpaceshipNewsNetw0rk"

Untrustworthy sources (news source):

- "SusAmongUs on reddit",
- "RedNewsMedia.blogsite.web",

- "Innocent Street Journal",
- "Spacehip News Network",
- "Imposter Kyle's Personal Blog",
- "Amogus Weekly",
- "Bluffington Post",
- "@GLSNewsBot",
- "The Gopher Report",
- "Vincenzo Moriarti, Amogus Weekly Editor-in-Chief",
- "Bobby Doubt, Host of The Gopher Report",
- "@CelebNewsBot",
- "@SusBot",
- "@SpaceshipNewsNetw0rk",
- "@BüffingtonPośt_"

Trustworthy sources (news source):

- "Spaceship News Network",
- "Amongus Weekly",
- "Buffington Post",
- "Crewmate Associated Press",
- "Zot Chronicles",
- "Julia Minnetian, Spaceship News Network Journalist",
- "Crew News",
- "Versace Pelligrini | Reporter @BuffingtonPost",
- "@JMinnetian | Journalist @SpaceshipNewsNetwork",
- "@SpaceshipNewsNetwork",
- "@VersacePelligrini",
- "@BuffingtonPost"