USENIX

THE ADVANCED COMPUTING
SYSTEMS ASSOCIATION

SoK: Come Together – Unifying Security, Information Theory, and Cognition for a Mixed Reality Deception Attack Ontology & Analysis Framework

Ali Teymourian and Andrew M. Webb, *Division of Computer Science & Engineering, Louisiana State University;* Taha Gharaibeh, *Division of Computer Science & Engineering, Baggil(i) Truth (BiT) Lab, Center for Computation and Technology, Louisiana State University;* Arushi Ghildiyal, *Division of Computer Science & Engineering, Louisiana State University;* Ibrahim Baggili, *Division of Computer Science & Engineering, Baggil(i) Truth (BiT) Lab, Center for Computation and Technology, Louisiana State University*

https://www.usenix.org/conference/usenixsecurity25/presentation/teymourian

This paper is included in the Proceedings of the
34th USENIX Security Symposium.

August 13–15, 2025 • Seattle, WA, USA

978-1-939133-52-6

Open access to the Proceedings of the
34th USENIX Security Symposium is sponsored by USENIX.

SoK: Come Together – Unifying Security, Information Theory, and Cognition for a Mixed Reality Deception Attack Ontology & Analysis Framework

Ali Teymourian[1], Andrew M. Webb[1], Taha Gharaibeh[1,2], Arushi Ghildiyal[1], Ibrahim Baggili[1,2]
[1]*Division of Computer Science & Engineering*
[2]*Baggil(i) Truth (BiT) Lab, Center for Computation and Technology*
Louisiana State University
{ateymo1, andrewwebb, tghara1, aghild1, ibaggili}@lsu.edu

Abstract

We present a primary attack ontology and analysis framework for deception attacks in Mixed Reality (MR). This is achieved through multidisciplinary Systematization of Knowledge (SoK), integrating concepts from MR security, information theory, and cognition. While MR grows in popularity, it presents many cybersecurity challenges, particularly concerning deception attacks and their effects on humans. In this paper, we use the Borden-Kopp model of deception to develop a comprehensive ontology of MR deception attacks. Further, we derive two models to assess impact of MR deception attacks on information communication and decision-making. The first, an information-theoretic model, mathematically formalizes the effects of attacks on information communication. The second, a decision-making model, details the effects of attacks on interlaced cognitive processes. Using our ontology and models, we establish the MR Deception Analysis Framework (DAF) to assess the effects of MR deception attacks on information channels, perception, and attention. Our SoK uncovers five key findings for research and practice and identifies five research gaps to guide future work.

1 Introduction

Mixed Reality (MR) is reshaping how we perceive and interact with our physical surroundings. In 2023, the global MR market surged to $4.6 billion, fueled by leading tech giants Meta, Apple, and Microsoft [5]. MR headsets overlay virtual information onto the real world to assist human users, such as visualizing navigational aids on sidewalks to guide pedestrians. Malicious actors can exploit MR headsets to manipulate user perceptions and cause significant physical or social harm. For example, attackers can guide pedestrians into traffic by obstructing their view of oncoming vehicles.

Deception attacks pose a fundamental security threat for technologies that alter human perception of the real world. Deceptions introduce false beliefs or interpretations in a target [44]. Illusions, central to deception, lead to false perceptions of sensory input [46], achieved through deceit, where truthful information is hidden or false information is shown [2]. Using MR, attackers can affect information communication and decision-making, such as by introducing illusions (e.g., fake pedestrian crossings) or hiding essential information (e.g., navigation arrows). Protecting MR users is vital, yet we lack theoretical framing to describe and analyze MR deception attacks and their effects on human cognition.

This paper systematizes knowledge from disparate domains, introducing a framework for evaluating MR deception attacks. We address the following research questions:

RQ1: How does existing literature categorize MR deception attacks?

RQ2: How do we model the effects of MR deception attacks on information communication?

RQ3: How are the cognitive processes associated with decision-making affected by MR deception attacks?

RQ4: How can we systematically analyze MR deception attacks and their effects?

Our multi-stage methodology synthesizes knowledge from MR security, information theory, and cognition to derive our MR Deception Analysis Framework (DAF). First, we derived an MR deception attack ontology from the literature. Then, we integrated our ontology, an information-theoretic model of communication, and a cognitive decision-making model to derive our framework. Our work contributes the following:

- the **first in-depth investigation of deception attacks in MR environments**;
- a **deception analysis framework for assessing the effects of MR deception attacks** on information channels and decision-making;
- an **ontology of MR deception attacks**;
- an **information-theoretic model of MR deception attacks** that formalizes effects on communication;

- a **decision-making model of MR deception attacks** that connects cognitive processes, attacks, and effects;
- a **literature review of deception attacks** in MR;
- an **assessment of state-of-the-art MR technical attacks** use or potential use in deception attacks.

This paper is structured as follows. Section 2 grounds our work in foundational research. Section 3 outlines our methodology. Section 4 presents a literature review of MR deception attacks. Section 5 describes an ontology of existing attacks. Sections 6 and 7 develop information-theoretic and decision-making models to assess how MR deception attacks affect communication and cognition, respectively. Section 8 introduces our MR Deception Analysis Framework. Section 9 discusses implications and limitations of this work. Section 10 summarizes our contributions and suggests future work.

2 Background

We ground this work in foundational research on deception, information processing, decision-making, and MR.

2.1 Deception

Deception entails intentional acts to cultivate a belief in a recipient that the deceiver considers false [70, 105]. In order to induce false beliefs, communication is required [69]. This communication may be verbal or nonverbal. Deception can be modeled as information processing where a sender presents "truthful or false information (a signal) to an opponent (the receiver) in order to gain an advantage over the opponent" [23]. Separate cognitive processes exist for sender (deceiver) and receiver [47]. Accounts of deception must consider how "information sharing is dominated by unstructured communication involving natural language and a diverse collection of nonverbal cues" [47]. MR is primarily a visual medium where deceptions will often rely on nonverbal stimuli.

Models of deception center around interpersonal communication [16, 36, 37, 49, 58] or information transmission [11, 50, 64, 66]. The Interpersonal Deception Theory (IDT) [16] examines deception as an interactive, reciprocal relationship where both senders and receivers adapt their strategies in real-time. IDT integrates cognitive and emotional dimensions, such as arousal and suspicion, which influence deceptive behaviors and detection mechanisms during interpersonal exchanges. Levine's Truth-Default Theory [58] identifies cognitive biases underlying deception detection and shows that humans generally operate under a presumption of honesty. This facilitates efficient communication but leaves individuals vulnerable to deceit. The Emotion Deception Model [36,37] considers how both current emotions and anticipated emotions influence decisions to use deceptions during negotiations. McCornack [64] models deception as manipulations of information, emphasizing how individuals exploit conversational norms to mislead others while maintaining an appearance of cooperative communication. Borden [11] and Kopp [50] separately proposed models of deception that are grounded in information theory. The Borden-Kopp model categorizes four deception strategies for manipulating a victim's perception: Degradation (conceal information), Denial (increase uncertainty), Corruption (create false belief), and Subversion (alter information processing). We use these strategies as the basis for the foundational organization of an MR deception attack ontology and analysis framework.

2.2 Information Processing

Information processing theory emerged as a way of understanding human cognition, particularly problem-solving and decision-making, alongside advancements in computing during the 1950s and 1960s [87]. In this theory, computational models describe how humans acquire, process, and store information to make decisions and take actions. The information processing model operates in a serial manner. First, information is input through sensory receptors in the body. Then, information is sequentially stored in working (short-term) memory and mentally processed in decision-making. Finally, responses are output as human actions. In order to avoid sensory overload, attention mechanisms filter what information is stored and processed. We use information processing theory to derive our MR Deception Decision-Making Model (Section 7), which connects sensory input transmitted from MR headsets to attention, memory, and other cognitive processes.

2.3 Decision-Making

Decision-making is a complex cognitive process that is susceptible to deception [30]. It consists of three stages [33]. First, sensory input is processed to make assessments and predictions on possible outcomes. Second, cognitive processes select an action based on the perception of inputs and predictions of outcomes. Third, action responses are assessed to evaluate the outcome. Individuals often do not evaluate risks based on mathematical probabilities [48]. Instead, psychological factors, such as the certainty effect, play crucial roles. With the certainty effect, humans give more weight to outcomes that are seen as certain compared to those that are merely probable. This insight is valuable when anticipating MR user responses to deceptive stimuli, where the perception of risk and reward can be manipulated. Niforatos et al. [73] point out the complexities of ethical decision-making within MR, emphasizing the impact that immersive technologies have on human cognitive evaluations.

2.4 Mixed Reality

Milgram and Kishino [68] defined MR as a continuum of blended visual representations residing between the entirely

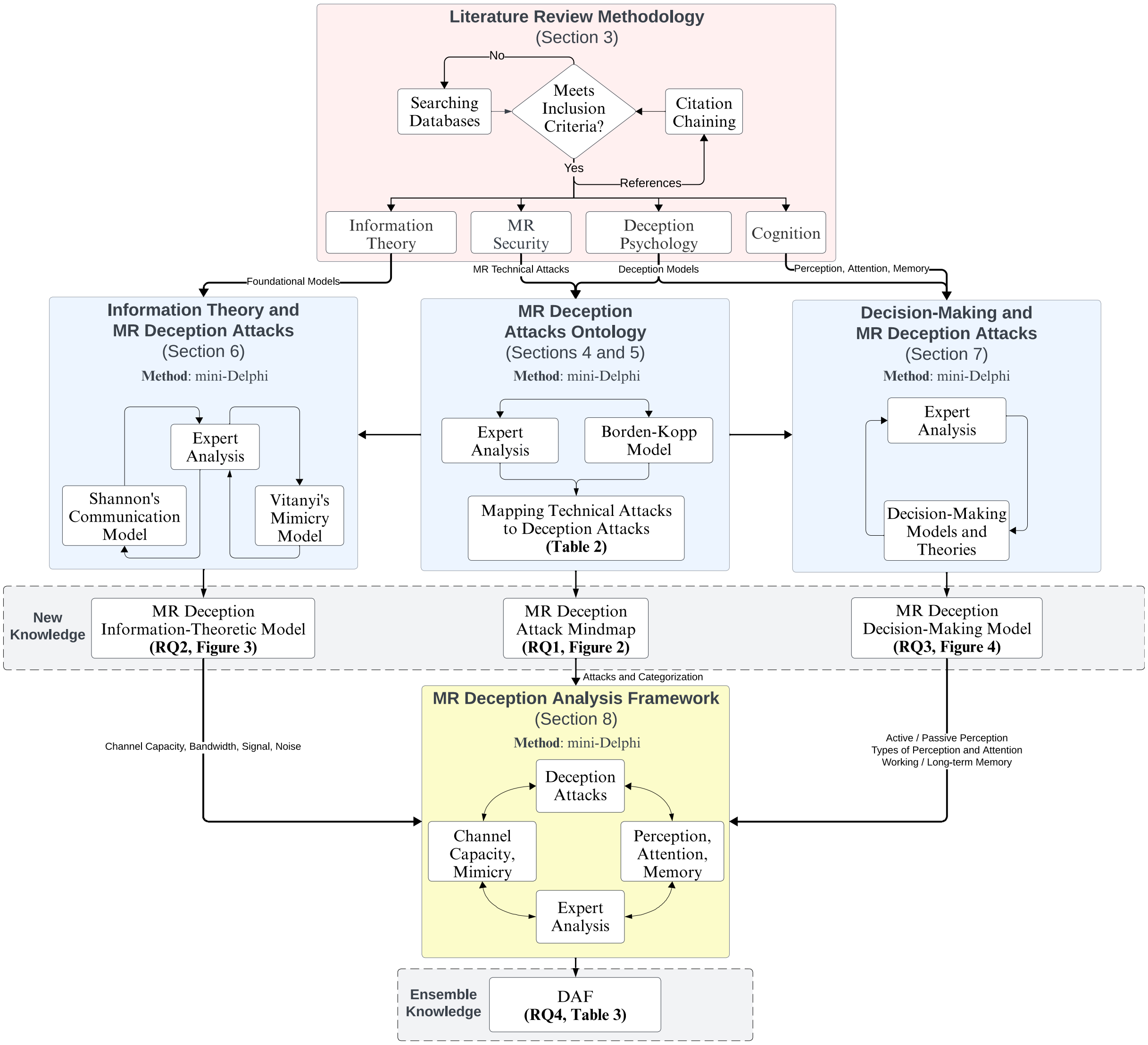
Literature Review Methodology
(Section 3)
No
Searching Databases
Meets Inclusion Criteria?
Citation Chaining
Yes
References
Information Theory
MR Security
Deception Psychology
Cognition
Foundational Models
MR Technical Attacks
Deception Models
Perception, Attention, Memory
Information Theory and MR Deception Attacks
(Section 6)
Method: mini-Delphi
Expert Analysis
Shannon's Communication Model
Vitanyi's Mimicry Model
MR Deception Attacks Ontology
(Sections 4 and 5)
Method: mini-Delphi
Expert Analysis
Borden-Kopp Model
Mapping Technical Attacks to Deception Attacks (Table 2)
Decision-Making and MR Deception Attacks
(Section 7)
Method: mini-Delphi
Expert Analysis
Decision-Making Models and Theories
New Knowledge
MR Deception Information-Theoretic Model (RQ2, Figure 3)
MR Deception Attack Mindmap (RQ1, Figure 2)
MR Deception Decision-Making Model (RQ3, Figure 4)
Attacks and Categorization
Channel Capacity, Bandwidth, Signal, Noise
Active / Passive Perception
Types of Perception and Attention
Working / Long-term Memory
MR Deception Analysis Framework
(Section 8)
Method: mini-Delphi
Deception Attacks
Channel Capacity, Mimicry
Perception, Attention, Memory
Expert Analysis
Ensemble Knowledge
DAF (RQ4, Table 3)

Figure 1: Our five-stage methodology beginning with literature review (top). Outcomes of the literature review informed intermediary stages. Knowledge from these stages culminates in the MR Deception Analysis Framework (DAF).

real and the fully virtual. Within their continuum are two forms of MR: Augmented Reality (AR) and Augmented Virtuality (AV). In AR, virtual elements overlay physical reality. Virtual Reality (VR) headsets, such as the Apple Vision Pro or Meta Quest 3, now support AR experiences through video pass-through where virtual information is overlaid onto camera feeds of the real world. In contrast, AV integrates real-world elements into VR. For example, many VR headsets display boundaries of physical spaces as users approach to avoid collisions. In this work, we focus on AR and AV systems that facilitate complex information processing scenarios. This complexity raises questions about how users interact with information in MR and the potential cognitive risks or vulnerabilities in decision-making. To the best of our knowledge, this is the first work exploring the impact of MR deception attacks on information communication and decision-making.

3 Methodology

We employed a systematic methodology that included an extensive literature review and development of theoretical models to describe the MR deception attacks. The literature review identified relevant theories, models, attacks, and empirical outcomes, which informed each step in our process (Figure 1). Using our review and expert knowledge, we derived an ontology of MR deception attacks (Figure 2). We connected technical attacks from the literature to our ontology (Table 2). Subsequently, we integrated an information-theoretic perspective to examine how deception attacks impact information communication in MR (Figure 3). Next, we developed a decision-making model that describes how cognitive processes handle sensory stimuli from MR headsets (Figure 4). Finally, we combined our ontology and two mod-

Table 1: Reviewed Articles Classified by Research Area.

Categories	Articles	#
MR Security		
User Manipulation & Deception	[4, 13, 17, 20, 22, 25, 56, 57, 74, 94–96, 100, 102]	14
Privacy and Data Security	[3,18,21,34,53,54,61,62,65,72, 75,81,86,88,89,93,99,103,104]	19
Frameworks / Surveys	[1,35,41,42,82,91]	6
Information Theory	[15,16,52,60,85]	5
Cognition		
Perception	[19,28,38,79,80,101]	6
Attention	[43,45,63,71,77,78,90,92]	8
Memory	[6,8–10,29,32,83,98]	8
Deception Psych.	[23,27,31,36,37,39,44,47,49, 59,66,69,70,97]	14
	Σ Total:	80

els to derive a framework for assessing the cognitive effects of MR deception attacks on decision-making (Table 3).

Literature Review (Section 4): We conducted a systematic literature review covering a wide range of topics, including deception, privacy, perceptual manipulation, cognition, and decision-making. We used Google Scholar, ACM Digital Library, IEEE Xplore, MIT Press, and APA PsycArticles. Search terms included mixes of "AR/VR Security", "MR Deception", "Perceptual Manipulation", and "Decision-Making". We collected articles from reputable journals and conference proceedings, including USENIX Security, S&P, ISMAR, IEEE VR, and the Journal of Experimental Psychology. We limited articles to those published in the past five years to ensure relevance to current MR technologies. However, we additionally took into account important historical works that had significant impact. We focused on articles with well-defined research questions, comprehensive analysis, and innovative contributions.

The filtering process began by retrieving over 200 articles from search engines and databases. Two researchers screened titles and abstracts for relevance to ensure a consensus-based approach. The criteria for relevance included: alignment with **MR security**, **information theory**, **cognition**, and **deception psychology**; presence of well-defined research questions; and contributions to the field's innovation and depth of analysis. We reviewed full texts to confirm suitability based on the depth of analysis, innovation, and relevance to our research objectives. Articles were excluded if they lacked depth of analysis, innovation, or relevance to the key themes. This resulted in a final selection of ($n = 80$) articles across different domains, which are categorized in Table 1.

MR Deception Attacks Ontology (Section 5): After our literature review, two researchers iteratively outlined an encyclopedic map of identified deception attacks. The iterative process was enhanced using the mini-Delphi method [24, 76] in which two subject matter experts reviewed and further refined the ontology across each iteration. The outcome was a mind map of deception attacks in MR (Figure 2). Then, we characterized how technical attacks identified in our literature review fit within the newly developed ontology (Table 2).

Information Theory and Deception Attacks (Section 6): We derive an information-theoretic model to describe how MR deception attacks affect information communication. We employ Borden-Kopp's deception model [52] that uses Shannon's information theory [85] to formulate how information is transmitted from a source (e.g., MR application) to a user via a MR headset. Additionally, we utilize Vitanyi's model of mimicry [60] to mathematically assess differences between source-generated messages and those created by an attacker.

Decision-Making and Deception Attacks (Section 7): We used our MR Deception Ontology and our review of deception psychology and cognition literature to develop an MR Decision-Making Model. This model connects cognitive processes of perception, attention, memory, and decision-making to types of deception attacks. This process involved a mini-Delphi approach in which renditions of the model were iteratively revised using expert knowledge and prior literature.

MR Deception Analysis Framework (DAF) (Section 8): We utilize our two models to develop a framework for analyzing the effects of MR deception attacks on information communication and human cognition. This framework allows for qualitatively evaluating the impact of MR deception attacks on cognitive processes associated with perception, attention, memory.

> **Key Finding (KF) 1.** Our multidisciplinary methodology shows how to connect disparate knowledge into an ensemble framework. As computing becomes more ubiquitous, security challenges require broader perspectives and analysis, particularly in terms of human cognition. We are not aware of other work that connects literature and theories from cybersecurity, MR, and cognitive sciences into a cohesive framework.

4 MR Attacks and Surveys

Our literature review categorizes unique aspects of MR security into three distinct areas: User Manipulation and Deception, Privacy and Data Security, and Frameworks and Surveys.

4.1 User Manipulation and Deception

Prior work explored techniques to manipulate facets of users' perceptions and decision-making in MR. Casey et al. [17] introduced new proof-of-concept attacks that pose a threat to user safety in a Virtual Environment (VE). Their work categorized and defined the following attack types: chaperone, disorientation, human joystick, and overlay. The human joystick successfully manipulated users to move to specific physical locations without their awareness. The chaperone attacks manipulated the VE boundaries, while the disorientation attack elicited a sense of dizziness and confusion from an immersed VR user. Lastly, in an overlay attack, an adversary overlaid objects such as images and videos onto a user's VR view. Chandio et al. [20] introduced stealthy and practical multi-modal attacks on MR tracking, showing that MR systems relying on sensor fusion algorithms for tracking can be compromised through perceptual manipulation by attacking multiple sensing streams simultaneously.

Nilsson et al. [74] provided an overview of Redirected Walking (RDW) techniques in VR that use subtle manipulations of gains and overt redirection techniques to manipulate user's perception of space and movement. Brinkman [13] describes attacks that subtly influence user choices without their awareness as decisional interference. De Haas & Lee [25] provide a comprehensive analysis of the manipulative potential of audio effects design in AR which systematically categorizes deceptive audio cues into various categories, each of which uniquely influences user perception and behavior. Wang et al. [100] further investigate how these deceptive design techniques, known as dark patterns, can manipulate users in AR environments and compromise their information and safety. Building on psychological aspects of manipulation, Trivers [94] describes how deception is a natural part of life, not just for humans but all living beings. This analysis provides a foundational understanding of the psychological dynamics at play, illustrating how MR systems can exploit the natural tendencies of humans to manipulate and be manipulated, influencing user perception and decision-making.

Perceptual Manipulation Attacks (PMAs) attempt to exploit a user's sensory perceptions to influence their decision-making, which can lead to physical harm [22,95]. Ali et al. [4] investigated visual deception by creating illusions of 3D views using projections onto 2D surfaces. Cheng et al. [22] derived a framework for comprehending and addressing PMAs within the context of MR. They demonstrated that PMAs can manipulate user perceptions to affect reaction times. They investigated the effects of PMAs on situational awareness, revealing how MR content can divert users' attention away from essential real-world stimuli, undermining their concentration and attentiveness. Ledoux et al. [56] found that visual cues in VR can evoke food cravings, showing how sensory manipulations influence user perception. Tseng et al. [95] investigated the risks of perceptual manipulations in VR, focusing on the negative impacts that these manipulations may have on users.

Research Gap (RG) 1. Most recent research on user manipulation and deception has focused on VR systems, leaving AR systems underrepresented. Future research should prioritize AR security.

4.2 Privacy and Security

MR and VR headsets pose significant challenges for privacy and security. These headsets collect, use, and present personal information, making them vulnerable to information leaks via side-channel attacks. Further, attackers can use deception attacks to disrupt information channels and cognitive processes causing users to take actions that may expose additional personal information.

Slocum et al. [88] introduced TyPose, which uses machine learning techniques to classify motion signals from MR headsets by analyzing subtle head movements made by users when interacting with virtual keyboards. Al Arafat et al. [3] presented the VR-Spy system, which utilizes the channel state information of Wi-Fi signals to detect and recognize keystrokes based on fine-grained hand movements. Su et al. [93] present a method for remotely extracting motion data from network packets and correlating them with keystroke entries to obtain user-typed data such as passwords or private conversations. Ling et al. [61] highlighted the vulnerability of VR systems to novel side-channel attacks. They showed how these attacks exploit computer vision and motion sensor data to infer keystrokes in a VE. Knowing what information a user is typing or specific personal details could help attackers develop more believable deception attacks.

Vondráček et al. [99] introduced the Man-in-the-Room attack in VR, where an attacker gains unauthorized access to a private VR room and observes all interactions. Through observation, attackers can develop more targeted deception attacks. Nair et al. [72] outlined significant privacy risks in VR environments, proposing a threat model with four adversaries: Hardware, Client, Server, and User. These adversaries have access to different aspects of the VR information flow. These risks can covertly reveal personal data, and adversarially designed VR games may manipulate users into disclosing sensitive information.

Prior work has also explored the digital forensics of VR headsets. Yarramreddy et al. [103] presented an exploration of the forensics of immersive VR systems, which demonstrates the feasibility of reconstructing forensically relevant data from both network traffic and the systems themselves. Casey et al. [18] introduced the first open-source VR memory forensics plugin for the Volatility Framework. Using forensic techniques could allow an attacker to uncover personal information about a user's behavior or interest, which could be leveraged for deception attacks.

Security issues can expose MR users to physical harm and potential deception attacks. Odeleye et al. [75] showed attacks targeting GPU and network vulnerabilities in VR systems to manipulate frame rates and cause VR sickness. Roesner et al. [81] conducted a comprehensive examination of security and privacy concerns in AR, unveiling new vulnerabilities unique to AR applications. For example, they suggest displaying the provenance of AR elements so that users know the source of augmentations. Without this, users are susceptible to deception attacks that inject false information. McPherson et al. [65] conducted the first system-level assessment of security and privacy features in AR browsers. Lebeck et al. [53] introduced Arya, an AR platform aimed at regulating application output to mitigate risks from malicious or faulty applications. This focus on output security is complemented by research delving into input privacy risks and the largely unexplored area of malicious AR output [54]. Cheng et al. [21] introduced several proof-of-concept attacks targeting User Interface (UI) security vulnerabilities in AR systems. Slocum et al. [89] investigate the security vulnerabilities in multi-user AR applications, focusing on the shared state that maintains a consistent virtual environment across users.

4.3 Frameworks and Surveys

Garrido et al. [35] systematized knowledge on VR privacy threats and countermeasures, focusing on two types of attacks: profiling and identification. Profiling attacks collect sensitive personal data to create user profiles. Identification attacks uniquely pinpoint a user within a VR environment. Happa et al. [42] developed an abstraction-based reasoning framework to reveal possible attacks in collaborative MR applications. De Guzman et al. [41] provided a survey of various protection mechanisms proposed for MR. Adams et al. [1] conducted interviews with MR users and developers to survey MR privacy policies and their perceptions. Stephenson et al. [91] systematized knowledge on AR/VR authentication mechanisms, evaluating research proposals and practical deployments.

> **RG 2.** There is a notable lack of frameworks that address diverse aspects of MR security, including technical exploits, user experience, detection, and defense.

5 MR Deception Attacks Ontology

We derive a MR deception attack ontology (Figure 2) from our review of the literature, our expert knowledge, and the Borden-Kopp model [11, 15, 50]. Their model focuses on how deceptions alter a victim's decision-making by manipulating information channels to inject false information or hide true information. Additionally, they identify how false information can be used to induce biases that influence how information is processed and interpreted. As MR headsets directly transmit information to users, their information-theoretic model provides an appropriate and robust framework for describing, categorizing, and analyzing MR deception attacks. The Borden-Kopp model divides deception attacks into *channel attacks* and *processing attacks*.

5.1 Channel Attacks

Channel attacks primarily target information communication channels. These attacks exploit the physical or logical paths that data takes as it moves between different components of a system or between different entities. The Borden-Kopp model identifies three types of channel attacks: *overt degradation*, *covert degradation*, and *denial*.

5.1.1 Overt Degradation

With overt degradation, attackers of MR systems create confusion by introducing substantial visual, auditory, or tactile noise to prevent victims from accurately perceiving or engaging with virtual objects, the physical world, or associated tasks. Due to the blatant nature of overt degradation, victims become aware that they are under attack. The presence of virtual noise can be disorienting in the context of MR, as users heavily depend on the seamless integration of real and virtual information in order to maintain focus on a task. Further, it can disrupt immersive experiences, preventing users from becoming fully engaged in a task. We identify the following forms of overt degradation attacks:

- *Sensory Overload*: Inundate the user's sensory receptors with excessive amounts of stimuli, leading to disorientation or distraction [75, 81]. Disorientation can cause a user to feel lost or confused within a VE, making them more susceptible to manipulation. Distraction diverts the user's attention, potentially preventing them from detecting or responding to an attack.
- *Momentary Misdirection*: Redirect the user's attention using virtual content within a MR systems. Misdirection distracts the user from their task. For example, an attacker can insert flashing virtual elements that draw the user's visual attention away from seeing important information or activities in the physical world.
- *Signal Replacement*: Alter or replace sensory input within MR systems. This can lead to a user perceiving a different reality from what actually exists, potentially causing confusion, disorientation, or exploitation [95].
- *Quality Erosion*: Reduce the quality of the signal from the MR headset. This can be achieved through actions such as decreasing the resolution of visual elements, introducing distortions to audio, or reducing the vibration intensity of haptic feedback.

5.1.2 Covert Degradation

Covert degradation attacks subtly suppress or diminish the clarity of information presented by MR headsets. Attackers can blend deceptions seamlessly with the MR environment, thereby making it harder for the user to discern. By leveraging immersive MR experiences, deception attacks can mask

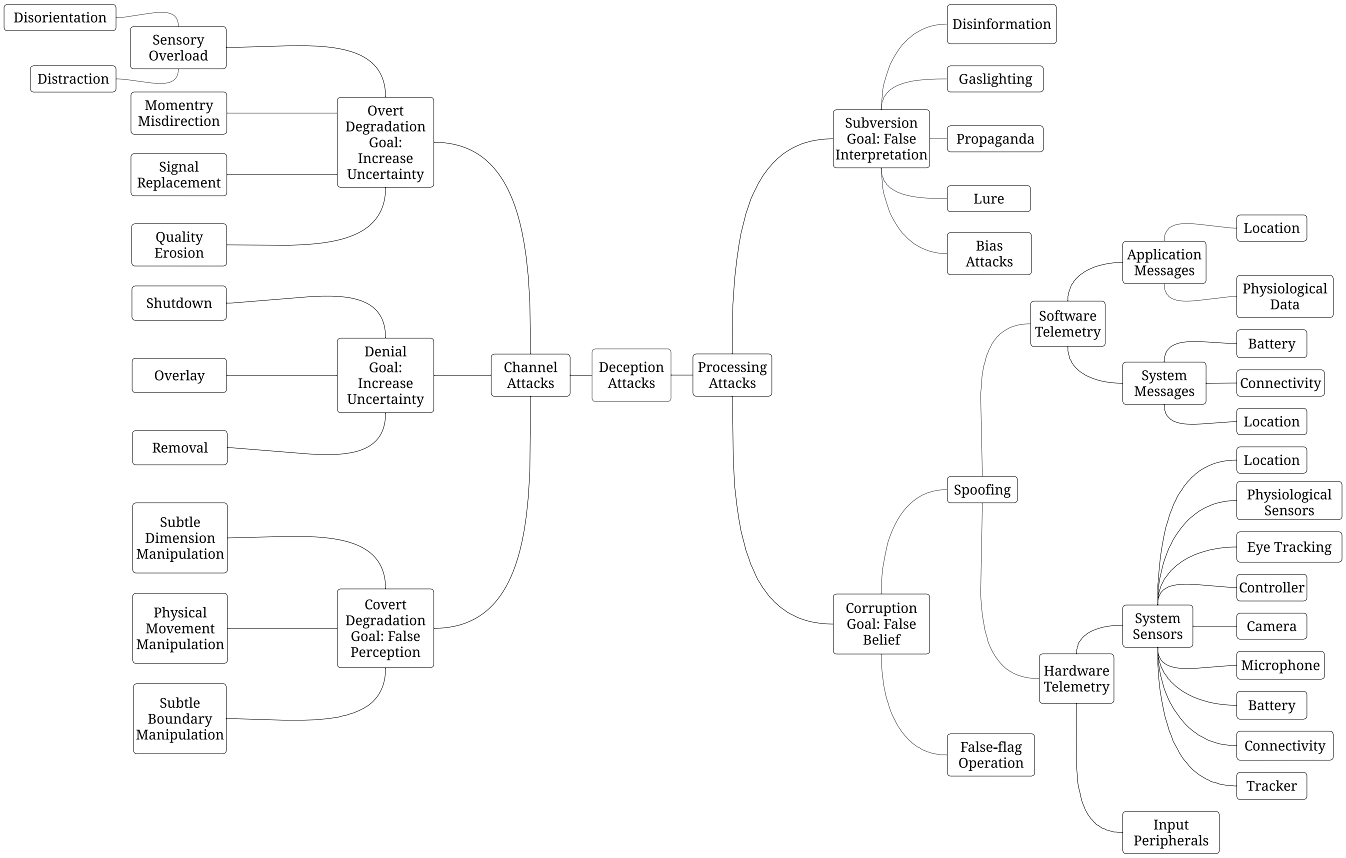

Figure 2: Mind Map of MR Deception Attacks Ontology. Channel attacks on the left. Processing attacks on the right.

false information as users' attention and interactions are focused elsewhere. We identify the following forms of covert degradation attacks:

- *Physical Movement Manipulation*: Relocate a user without their awareness or consent by discreetly shifting the center of a VE while they focus on a task [17].
- *Boundary Manipulation*: Altering boundaries within the VE, which can lead to unexpected collisions with objects or distortions in spatial perception [84].
- *Dimension Manipulation*: Modifying the proportions, scale, or spatial relationships of virtual objects [12].

5.1.3 Denial

Denial attacks seek to increase uncertainty by obstructing the user's access to information. This is achieved by shutting down virtual overlays, prohibiting interaction with virtual objects, or disrupting the seamless blend of real and virtual elements. This is often an overt method of deception, as users may be cognizant of their deprived or diminished accesses [51]. A user may find themselves subject to a Denial attack if they lose ingress to existing networks, communication channels, and various other system features. We identify the following forms of denial attacks:

- *Shutdown*: Deliberately terminate or disable a MR communication channel or service.
- *Overlay*: Layer content over a communication channel to disrupt normal operations of the channel [57, 81, 95].
- *Removal*: Selectively remove or block information [75].

5.2 Processing Attacks

Processing attacks target vulnerabilities in how humans cognitively process information, aiming to deceive humans by altering their perceptions, interpretations, and understandings of information. The Borden-Kopp model identifies two types of processing attacks: Corruption and Subversion.

5.2.1 Corruption

Corruption attacks deliberately manipulate the MR system by counterfeiting existing virtual elements and information. These manipulations result in inconspicuous data and actions that are difficult to discern from standard data and actions within the MR system. Their primary objective is to create

Table 2: Connecting Technical Attacks to MR Deception Attacks.

	Sensory Modality			Technical Modality					Channel Attacks										Processing Attacks						
									OD				CD			Denial			Corr.		Subversion				
Technical Attacks	**Visual**	**Auditory**	**Tactile**	**Hardware**	**Software**	**Network**	**Data**	**Side-Channel**	**Sensory Overload**	**Momentary Misdirection**	**Signal Replacement**	**Quality Erosion**	**Subtle Boundary Manip.**	**Subtle Dimension Manip.**	**Physical Movement Manip.**	**Shutdown**	**Overlay**	**Removal**	**Spoofing**	**False-Flag Operations**	**Bias Attacks**	**Lure**	**Disinformation**	**Propaganda**	**Gaslighting**
GPU-Based [75]	✓				■				❍			●				●									
Network-Based [75]	✓	✓	✓			■						●				❍									
Color [22]	✓				■												●								
Auditory [22]		✓			■				●	●															
Puppetry [95]	✓		✓		■		❑								●							●			
Mismatching [95]	✓				■		❑							●			●	●					❍		❍
Object-in-the-Middle [21]	✓				■									❍			●								
Object Erasure [21]	✓				■								❍	❍				●							
Chaperone [17]	✓				■	❑							●									❍	❍		
Human-Joystick [17]	✓				■	■	■								●							❍			
Inception [102]	✓				■	❑	■								●				❍				❍		
Man in the Room [99]	✓				■	■	■	❑											❍				❍		❍
Output Manipulation [81]	✓				■												●						❍		
Clickjacking [81]	✓				■					❍							●								
Cursor-Jacking [57]	✓				■						●						●								
Blind Spot [57]	✓				■						❍		●				●								
Read [89]	✓				■		■				●						●						❍		
Write [89]	✓				■		■				●						●		●	❍					
Hand Gestures Inference [104]	✓							■		❍							●								
Face-Mic [86]		✓		❑				■			❍								❍	❍					
TyPose [88]			✓					■			❍								❍	❍					
VRSpy [3]	✓							■			❍								❍	❍					
Remote Keylogging [93]	✓					❑		■											❍						
Heimdall [62]		✓						■			❍								❍						
LocIn [34]	✓				■		■								❍				❍						
Zero Displacement [20]	✓			■	■		■	■		❍				●									❍		
Speed Manipulation [20]	✓			■	■		■	■							●				●		❍				
Path Deviation [20]	✓			■	■		■	■			●		●						●				❍		
Side-Swing and Switching [96]	✓	✓		■				■											●		●				
Fabrication of False Narratives [14]	✓				■														●				●		
Non-Verbal Manipulative Persuasion [14]	✓	✓			■						●													●	
Selective Exposure [14]	✓	✓			■						●										●				

■ Primary Technical Modality ❑ Secondary Technical Modality
● Mentioned in the article ❍ Not specifically mentioned, but can be deployed using the attack

false belief in a user, often causing compromised decision-making, incorrect conclusions, or virtual misdirection. Due to the immersiveness of MR, users may be more susceptible to corruption attacks as their engagement keeps them preoccupied, preventing critical analysis of false information. We identify the following corruption attacks:

- *Spoofing*: Create or modify data in a way that deceives the recipient or system into believing that the data is authentic or unaltered. Two forms of spoofed data are:

  - *Software Telemetry*: Alter or fabricate telemetry data from software. Attackers create or manipulate telemetry messages that convey a normally functioning application. Further, attackers may spoof telemetry messages at the system level, affecting multiple applications or impacting critical systems [20].
  - *Hardware Telemetry*: Alter or fabricate telemetry data from hardware sensors. Attackers can generate false sensor readings. Alternatively, attackers can manipulate input data from MR headsets or peripherals, such as controllers, enacting undesired actions or preventing users from performing desired tasks [20, 96].

- *False-Flag Operations*: Disguise the source of an attack in order to blame another party.

5.2.2 Subversion

Subversion attacks covertly manipulate a system or its information streams, resulting in falsified and fabricated interpretations by the user. Subversion often employs covert tactics, such as corruption attacks, which weaken trust or disrupt normal operations. We suspect that the immersiveness of MR can aid false interpretations as users unknowingly engage with deceptive information through repeated interactions, which can correspondingly build trust in deceptive elements. We identify the following subversion attacks:

- *Bias Attacks*: Deliberate manipulation of data or decision-making processes to systematically introduce bias or prejudice toward a specific concept or outcome.
- *Disinformation*: Spread false information to deceive and cause harm [40].
- *Lure*: Entice users to engage with (harmful) content.
- *Propaganda*: Manipulate perceptions, influence narratives, and garner support for a specific cause or element.
- *Gaslighting*: Erode trust and confidence, making it difficult for victims to distinguish truth from deception.

5.3 Connecting Technical Attacks to Ontology

MR deception attacks in our ontology typically rely on technical attacks to facilitate access to MR systems. Table 2 characterizes the modalities and deception attacks supported by each technical attack identified in our literature review. For each technical attack, we identify deception attacks directly mentioned by the authors (●) and deception attacks where the technical attack could be deployed but was not specifically mentioned by the authors (❍). We found more Channel Attacks (23) mentioned than Processing Attacks (8). This is not surprising considering that technical attacks typically target system-level functions which have more impact on the communication channels of MR headsets than user's cognitive processes. Still, we see seven attacks that mention Corruption or Subversion, and another eleven that we consider capable of supporting Processing Attacks.

> **RG 3.** State-of-the-art MR technical attacks predominately enable Channel Attacks. More research is needed on technical attacks that facilitate Processing Attacks and how these attacks affect MR users.

We identify the sensory modalities affected by an attack and the technical modalities it targets. Sensory modalities include visual, auditory, and tactile (e.g., vibrotactile feedback from controllers). Technical modalities include hardware, software, network, data, and side-channel [7].

> **KF 2.** Technical attacks primarily target the visual and software modalities. MR headsets include displays and processors, making visual and software modalities convenient targets. These attacks particularly focus on overlaying content or replacing signals as opposed to overloading, eroding, or removing signals. The least targeted modalities are tactile and hardware.

6 Information Theory and Deception Attacks

While our ontology categorizes MR deception attacks, it does not explore the effects of these attacks. To address **RQ2**, we use Kopp et al.'s framework [52], which connects Borden-Kopp's deception model [15] and Shannon's communication model [85], to derive an information-theoretic model of MR deception attacks. Shannon's communication model describes how information is transferred from a source to a destination as a message. The message is sent as a signal through a transmitter to a receiver. During transmission, the message is affected by noise, which combines with the signal. In our model, the transmitter is an MR headset, which acquires information from a source (e.g., an application, sensor, web service), and transmits that information in visual, auditory, or tactile forms to a user (destination) via displays, speakers, and controller vibrations (Figure 3). MR deception attacks affect the capacity of information transmission by introducing noise to degrade messages, denying access to information, or inserting fake information into messages.

6.1 Channel Capacity

According to Shannon's channel capacity theorem [85], the capacity of a channel to transmit information depends upon several factors, including the magnitude of the signal used to encode symbols, the level of interfering noise present in the channel, and the bandwidth of the channel.

$$C = W \log_2 \left(1 + \frac{S}{N_A + N_E}\right) \tag{1}$$

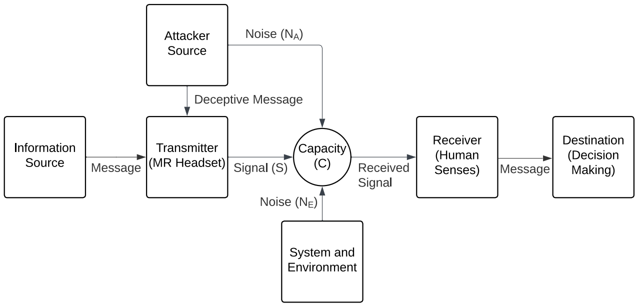

Figure 3: MR Deception Information-Theoretic Model. Messages are transmitted by a MR headset to a user. Deceptive messages are injected into transmissions. Noise from the attacker or environment affect channel capacity.

Channel capacity C represents the maximum amount of information that can be effectively transmitted from a source to a destination in bits per second (Equation 1). Bandwidth W refers to the information transfer rate of the communication channel in hertz. As W decreases, channel capacity correspondingly decreases through a linear relationship.

For MR headsets, information is transmitted through a headset to a human user. Thus, channel capacity determines how much visual, auditory, and tactile information can be transmitted. Signal S is the virtual content transmitted from the headset through displays, speakers, and vibrotactile motors. Noise N is categorized into two types: N_A which represents noise from an attacker source, and N_E, which represents noise from the real-world environment as well as noise that comes from the system itself, such as rendering stutters or audio glitches. N_A refers to potential external interference or malicious disruptions. N_E encompasses both ambient disturbances from the surrounding environment and internal system issues that can affect the MR experience. Both these sources of noise have a negative effect the channel capacity.

6.2 Channel Attacks

Channel attacks target channel capacity through reducing bandwidth, manipulating the signal, or introducing noise. Denial attacks involve an adversary's intention to significantly reduce access to the signal by primarily manipulating bandwidth. The channel capacity C tends to zero as the bandwidth W tends to zero. By shutting down the device, the attacker completely blocks the signal output, and bandwidth (W) reduces to zero. Attackers can occlude task-specific information with other content, effectively reducing bandwidth and interfering with task performance. A Removal attack selectively removes information from the signal, reducing bandwidth as less information is transmitted per a second.

In Overt Degradation attacks, the adversary can introduce substantial levels of noise into the channel, decreasing the Signal-to-Noise Ratio (SNR). As the SNR tends towards zero, channel capacity C decreases and eventually reaches zero. In this case, the user is bombarded with excessive noise, making it impossible to distinguish between the intended content and the attacker's noise. An example of this attack is sensory overload, where an attacker overwhelms the user by emitting excessive sensory stimuli through the MR headset, resulting in disorientation and discomfort.

In Covert Degradation attacks, an adversary can reduce the signal strength, which results in a decrease in the SNR. As the signal tends toward zero, SNR also tends toward zero, decreasing C towards zero as well. In MR headsets, these attacks can involve subtle manipulation of sensory cues presented to a user. Subtle boundary manipulation and subtle dimension manipulation are examples of these attacks. Through subtle manipulation of boundaries or the sizes of virtual objects, the attacker can deceive the user into thinking they are not moving [17] or make it harder to interact with virtual objects.

6.3 Processing Attacks

Processing attacks manipulate cognition through deceptive methods that mimic the MR system. We use Vitanyi's model [60] to formalize how deceptive information and messages created by an attacker, X, differ from actual information and messages created by an MR system, Y:

$$D(X,Y) = \frac{K(XY) - \min(K(X),K(Y))}{\max(K(X),K(Y))} \tag{2}$$

$$M(X,Y) = 1 - D(X,Y) \tag{3}$$

where D represents the measure of difference, M represents the measure of similarity or mimicry, and K is the editing function applied to X and Y.

Corruption attacks involve altering data during transmission. Vitanyi's difference measure $D(X,Y)$ quantifies the degree of alteration between the original message X and the corrupted message Y. In MR, corruption attacks might involve unauthorized changes to visual information, such as application and system messages, as well as sensory information, including camera, geolocation, and battery status (Figure 2). Subversion attacks, on the other hand, involve manipulating how users interpret information within an MR system. These attacks require repeated corruption or covert degradation attacks to reduce user's trust and understanding. Thus, M must remain close to 1 as the user has a greater chance of detecting deceptions through repeated exposure.

> **RG 4.** While Vitanyi's model formalizes mimicry, we lack models that effectively describe how processing attacks impact human behavior. Specialized domains, such as formal methods in human-computer interaction, could offer valuable insights.

7 Decision-Making and Deception Attacks

Beyond effects on information channels, we seek to model how MR deception attacks impact human cognition. To address **RQ3**, which concerns the interactions between decision-making and MR deception attacks, we conduct a thorough review of the cognition literature and develop a comprehensive decision-making model that outlines the stages of decision-making susceptible to these attacks. Figure 4 shows our MR Decision-Making Model. The model provides an overview of how sensory input is cognitively processed by a user to make decisions and where the different types of attacks affect decision making. Our decision-making model comprises of seven components: Sensory Inputs, Attention, Perception, Memory, Decision-Making, Decision Execution, and Responses.

- *Sensory Inputs*: Visual, Auditory, Smell, Taste, and Touch are the five different types of sensory inputs that can be affected by deception attacks.
- *Attention*: Initial stage where sensory information is gathered. Provides a gateway to perception.
- *Perception*: Sensory information gathered is processed and understood.
- *Memory*: Processed information is stored in working or long-term memory for future use and retrieval.
- *Decision-Making*: Determining a particular course of action predicated on perception.
- *Decision Execution*: Decisions are executed.
- *Responses*: Physiological, behavioral, or cognitive responses of executed decisions.

7.1 Perception

Perception refers to the cognitive process through which one comprehends sensory stimuli [79]. Wang et al. [101] define perception as “a set of internal sensational cognitive processes of the brain at the subconscious cognitive function layer that detects, relates, interprets, and searches internal cognitive information in the mind.” Perception is either active or passive. Active perception involves the intentional direction of attention towards environmental stimuli to extract information [38]. In contrast, passive perception occurs without deliberate effort; sensory information is received as presented [80].

Perception involves three stages [79]:

- *Selection:* Filter and select environmental stimuli from meaningful experiences.
- *Organization:* Structure and categorize the selected information, creating coherent and stable perceptions through grouping by proximity and similarity.
- *Interpretation:* Assign meaning to organized stimuli, with individuals’ cultural or experiential backgrounds leading to different understandings of the same stimuli.

In each stage of perception, MR deception attacks can target specific vulnerabilities. During selection, attacks can cause sensory overload or misdirect focus on irrelevant stimuli. In the organization step, attacks could involve boundary or dimension manipulation, affecting how stimulus are structured and grouped due to changes in proximity or scale. Propaganda or spoofing attacks can target interpretation, affecting the meaning assigned to stimuli that may seem wrong but is coming from a trusted source (e.g., the system or a collaborator). These potential attacks highlight the importance of the accuracy and reliability of perception in MR systems.

In addition to the conscious components of perception, subliminal inputs play an important role in how individuals interact with and understand MR environments. Cetnarski et al. [19] show that subliminal stimuli—information presented below the threshold of conscious awareness—can significantly influence decision-making processes in MR. This underscores the need to understand these subtle interactions that occur at the subconscious level of perception.

7.2 Attention

James [45] described attention as the cognitive process by which the mind selectively concentrates on a singular element from a variety of possible stimuli or thoughts, emphasizing its essential function in creating our conscious perception. The seminal work of Posner [77] introduced a framework for understanding the neural bases of attention and its various components and extending James’s initial descriptions into a more nuanced understanding of the brain’s attentional mechanisms. Building upon these early foundations, attention classification has expanded to include four types:

- *Selective:* Focusing on relevant information while suppressing irrelevant information [71, 92].
- *Divided:* Capacity to allocate cognitive resources to multiple stimuli simultaneously, enabling individuals to engage in concurrent activities [90]. Attended stimuli are from the same sensory modality or different ones [43].
- *Sustained:* Readiness to perceive and respond to stimuli over prolonged periods, often without conscious awareness of this vigilance [63].
- *Executive:* Regulates cognitive and emotional responses through management of other cognitive processes [78]. Aids orchestration of thought and emotion in alignment with goals and the dynamic demands of the environment.

Channel attacks primarily target Selective and Sustained attention. They manipulate the sensory channels through which users receive information, affecting their ability to focus on relevant stimuli or maintain attention over time. Selective attention is exploited by degrading the sensory inputs, making it harder for users to distinguish between relevant and irrelevant stimuli. As mentioned in Section 5, this happens in overt and

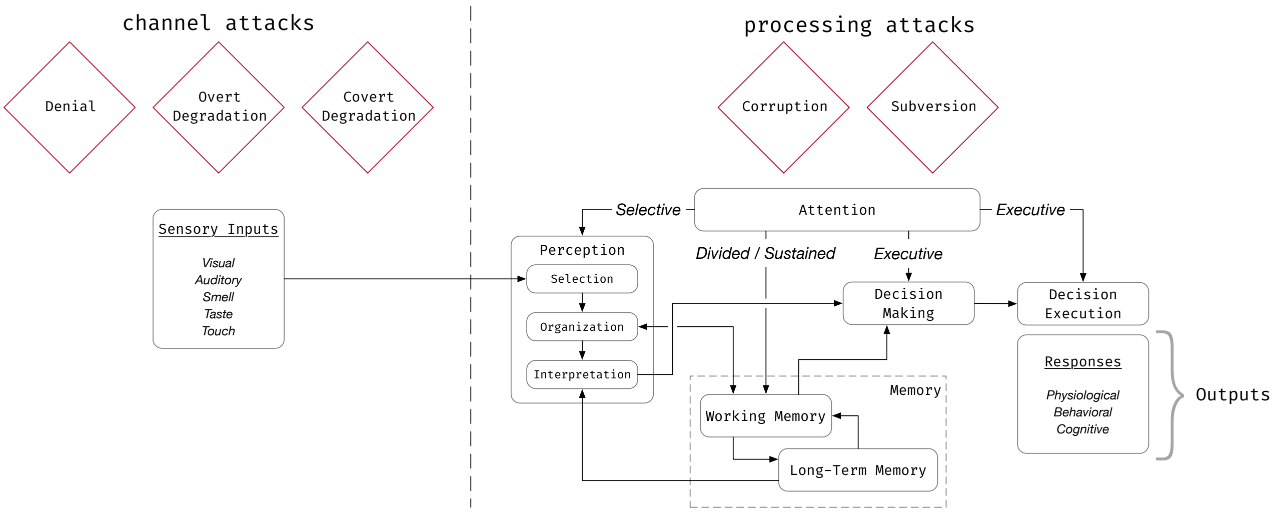

Figure 4: MR Deception Decision-Making Model. External stimuli (left) are input to cognitive processes (right). Stimuli are first processed by perception. Selective attention manages perception on relevant stimuli. Organized stimuli are stored in working memory. Interpreted stimuli are passed to decision-making, where executive attention manages decisions and their execution.

covert degradation attacks. These attacks may also limit sustained attention by making it more challenging for the user to maintain their focus over time, particularly when the quality of sensory inputs fluctuates or declines, resulting in increased cognitive load. Denial attacks block access to certain stimuli or information channels, disrupting selective attention.

Processing attacks primarily affect Divided and Executive attention by overloading the cognitive processing capabilities or by requiring constant adjustments to unexpected system behaviors. Corruption attacks can directly impact the users' selective and executive attention by altering the information presented within MR environment and also exploiting perceptual biases. Subversion attacks could challenge executive attention by forcing users to constantly adapt to unexpected system responses, requiring continuous updating of working memory. They also can target divided attention by interrupting the flow of tasks or actions within an MR environment, which compels users to divide their attention between correcting system errors and accomplishing their original goals.

> **KF 3.** Perception and attention are the primary targets for MR deception attacks. Channel attacks target selection mechanisms by degrading or denying stimuli. Processing attacks target interpretation and execution by corrupting beliefs or subverting interpretations.

7.3 Memory

Working memory and long-term memory are central components of our decision-making model. Baddeley [8–10] derived a multicomponent model of working memory consisting of the visuospatial sketchpad, phonological loop, central executive, and episodic buffer. The visuospatial sketchpad stores visual and spatial information while the phonological loop stores auditory and verbal information. The central executive directs attention towards stored information in either one. The episodic buffer provides temporary storage of information needed by the central executive with connections to the other three components and long-term memory. Long-term memory represents a permanent store that receives selected inputs from both the sensory register and working memory [6].

MR deception attacks affect memory and correspondingly attention. Downing [29] showed that the content of visuospatial sketchpad can guide selective attention toward matching visual stimuli. Through spoofing attacks, adversaries can produce deceptive stimuli that match expected stimuli, leveraging working memory to direct the user's selective attention. Santangelo and Macaluso [83] identified the critical role of working memory in managing divided attention when monitoring multiple objects simultaneously. Working memory load directly affects the efficiency of the central executive, with increased load impairing attention to multiple stimuli. Thus, sensory overload attacks can overwhelm working memory by visualizing too many objects for working memory to maintain. Unsworth & Robinson [98] suggested that individuals with lower Working Memory Capacity (WMC) may struggle more with maintaining consistent attention, leading to performance degradation in tasks requiring prolonged focus. Therefore, the impact of MR deception attacks that target WMC, such as sensory overload, will vary from person to person.

8 MR Deception Analysis Framework (DAF)

The culminating, ensemble knowledge that connects our ontology, information-theoretic model, and decision-making model

Table 3: MR attacks from our ontology are assessed according to the Information-Theoretic Model and Decision-Making Model.

		Models	Information-Theoretic Model				Decision-Making Model								
			C			*M*	Perception				Attention				Mem.
		Attacks	*W*	*S*	*N*		A/P	*Sel*	*Org*	*Int*	*Foc*	*Div*	*Sus*	*Exe*	
Channel Attacks	Overt Degradation	Sensory Overload			✓		A	███	███	███	███	███	███	███	W
		Momentary Misdirection			✓		A	▒▓█	▒▓█	▒▓█	███	▒▓█	▒▓█	▒▓█	W
		Signal Replacement		✓			A	███	███	███	███	███	███	███	W
		Quality Erosion	✓		✓		A	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	W
	Covert Degradation	Subtle Boundary Manipulation		✓			P	▒	▒	▒▓█	▒	▒	▒	▒▓	W
		Subtle Dimension Manipulation		✓			P	▒	▒▓	▒▓█	▒	▒	▒	▒▓	W
		Physical Movement Manipulation		✓			P	▒	▒	▒▓█	▒	▒	▒	▒▓█	W
	Denial	Shutdown	✓				A	███	███	███	███	███	███	███	W
		Overlay	✓		✓		A	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	W
		Removal	✓	✓			A	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	W
Processing Attacks	Corruption	Spoofing				✓	P	▒	▒	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	W
		False-Flag Operations				✓	P	▒	▒	▒▓█	▒	▒	▒	▒	W/L
	Subversion	Bias Attacks					P	▒	▒	▒▓█	▒	▒	▒	▒	W/L
		Lure					P	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	▒▓█	W/L
		Disinformation					P	▒	▒	▒▓█	▒	▒	▒	▒	W/L
		Propaganda					P	▒	▒	▒▓█	▒	▒	▒	▒	W/L
		Gaslighting					P	▒	▒	▒▓█	▒	▒	▒	▒▓█	W/L

Low = ▒, Low-Medium = ▒▓, Low-High = ▒▓█, High = ███
Information-Theoretic Model: C = Channel Capacity, W = Bandwidth, S = Signal, N = Noise, M = Mimicry
Perception: A/P = Active/Passive, *Sel* = Selection, *Org* = Organization, *Int* = Interpretation
Attention: *Foc* = Selective, *Div* = Divided, *Sus* = Sustained, *Exe* = Executive; **Memory**: W = Working, L = Long-Term

is the MR Deception Analysis Framework (DAF)—an assessment tool for identifying and discussing the multifaceted impact of MR deception attacks on user cognition (**RQ4**).

DAF classifies attacks according to their operational mechanisms, which can be overt or covert, involving Degradation, Denial, Corruption, or Subversion, and the cognitive processes they aim to disrupt. We focus on identifying where attacks manipulate MR communication channels by altering bandwidth (W), signal (S), noise (N), or by employing mimicry (M). Additionally, we explore the cognitive effects of each attack, examining the extent to which they can affect perception, attention, and memory. For perception and attention, we further breakdown analysis into stages of perception—Selection (*Sel*), Organization (*Org*), and Interpretation (*Int*)—and types of attention—Selective (*Sel*), Divided (*Div*), Sustained (*Sus*), and Executive (*Exe*).

Table 3 presents our general analysis of the different categories of attacks identified in our ontology. Overt Degradation and Denial attacks strongly affect both perception and attention. Covert Degradation, Corruption, and Subversion attacks primarily target the Interpretation stage of perception. These attacks typically require remaining hidden from the user. Thus, any effects on attention or early stages of perception are likely too revealing.

> **KF 4.** The interpretation stage of perception is a primary target of MR deception attacks. Deceptions seek to cultivate false beliefs, formed initially by interpretations of perceived stimuli.

For assessing the degree to which attacks affect stages of perception, we derived the following questions. Answers are either Low, Medium, High, or a combination of the three. De Meyer et al. [26] state that a three-point scale provides a practical balance between simplicity and reliability. It minimizes measurement error and ensures clarity in response, which can be important for consensus building in Delphi procedures.

- *Selection:* To what degree does the attack make it difficult to attend to or ignore task-related sensory stimuli from the physical or virtual environments during a decision-making task?
- *Organization:* To what degree does the attack make it difficult to group task-related sensory stimuli, such as by proximity or similarity, for a decision-making task?
- *Interpretation:* To what degree does the attack make it difficult to accurately assign meaning to organized, task-related stimuli and correctly interpret patterns and relationships within virtual and physical environments

when deriving understanding, making decisions, and taking action in a decision-making task?

For assessing the degree to which attacks affect types of attention, we derived the following questions. Answers are either Low, Medium, High, or a combination of the three.

- *Selective:* To what degree does the attack make it difficult to focus attention on relevant physical and virtual objects for a decision-making task in MR?
- *Divided:* To what degree does the attack make it difficult to switch between concurrent tasks rapidly while maintaining situational awareness in both the virtual and physical environments?
- *Sustained:* To what degree does the attack make it difficult to continuously scan and interpret information presented in the mixed reality environment, making timely decisions and adjustments?
- *Executive:* To what degree does the attack make it difficult to manage attentional resources effectively to interact with virtual elements while remaining aware of and responsive to the physical environment while performing a decision-making task?

DAF provides a systematic approach to evaluate threats posed by MR deception attacks. We posit that such analysis is pivotal for developing more resilient MR systems and training programs that can mitigate the impacts of deceptive threats.

KF 5. DAF is a tool for defining experimental research on MR deception attacks. We posit that it can be used to explore future attacks and may be extended for deception analysis beyond MR research.

RG 5. We need empirical findings to validate and precisely model the impact of MR deception attacks on cognitive processes and information channels.

9 Discussion

DAF provides a systematic method to classify and analyze MR deception attacks. While we focus on MR headsets, DAF is applicable to other forms of MR and even other areas of human-computer interaction (HCI). Kopp et al.'s information-theoretic framework [52] applied the Borden-Kopp model of deception to news media. We have broadened its use to MR deception attacks. Future work should extend the scope to other areas of HCI that involve information processing and decision-making. Our information-theoretic model and decision-making model are not tied to specific technologies or attacks, but rather provide generalizable models for studying the effects of deception in computing. To enhance DAF, future work should validate it empirically, expand its applicability to diverse contexts, incorporate individual cognitive factors, and refine models for processing attacks.

Researchers and practitioners can use DAF to assess the security threat of MR deception attacks. For example, we can assign values of 1 to 3 for Low to High ratings, respectively. Then, we can sum the values to identify which attacks pose the highest threat to perception and attention. Further, DAF can help develop deception detection and prevention approaches. For example, we can compare differences between rendered frames to see how the signal is changing. High volatility in changes may indicate overt degradation attacks, particularly if we can identify noise based on differences between expected and actual frames. More subtle changes that are spatial located in unexpected areas may indicate covert degradation attacks. Using eye-tracking sensors on these headsets, we can derive models of attention that can help identify when different types of attention are being employed or disrupted.

Limitations: This SoK synthesizes existing knowledge towards developing a field of study around MR deception. It is theoretical in nature and would benefit from further empirical validation. Controlled experiments involving MR deception attacks are essential for refining the framework and assessing its relevance to diverse scenarios. Furthermore, DAF does not fully account for cognitive diversity among users. Individual differences in cognitive capacity, attention, and susceptibility to deception are critical factors that could influence the effectiveness of both attacks and countermeasures.

10 Conclusion

MR technologies provide a wide range of opportunities while raising significant cybersecurity challenges. MR systems influence how we perceive physical and virtual environments, making them particularly susceptible to deception attacks. This multidisciplinary SoK brings together diverse knowledge to provide a systematic way for categorizing and analyzing MR deceptive attacks and their effects. It serves as a foundation and guide for future research. Our examination indicates that while there is a growing interest in MR security and a number of technical attacks, there is a lack of comprehensive research regarding deception attacks, particularly with regards to information communication and human cognition. Future work should investigate empirical studies of how MR deception attacks affect cognitive processes. Such studies can inform new techniques for detecting and mitigating threats from MR deception attacks. We envision DAF as a generalizable framework; however, specialized domains may possess nuances not fully captured by DAF. While we expect our information-theoretic approach to remain valid across MR technologies, where information is still transmitted and processed, it may be necessary to extend DAF through new metrics and additional models. We look forward to seeing how researchers can leverage DAF in studies of deception.

Acknowledgments

This work was supported by the Defense Advanced Research Projects Agency (DARPA), grant number HR00112320030. We appreciate the shepherd and anonymous reviewers for their insightful and valuable feedback.

Ethics Considerations

Conducting research on deception requires significant ethical considerations. At the forefront is mitigating risks to human participants and end users. When deceiving participants, it is necessary to ensure that benefits of the research far outweigh any potential risks to participants. Typically, research institutions have an IRB to enforce participant protections from unnecessary harm during human-subjects research. For MR research, harm can take many forms including physical, cognitive, technological, and social. As MR headsets affect how users perceive the physical world, deception attacks pose significant physical risk. Precautions must be taken to mitigate risks by screening out participants that may have adverse reactions to perceptual manipulations. Further, researchers should provide safe environments where participants cannot harm themselves by colliding with objects or falling down. Researchers should also consider how deceptive information may impact participants trust and understanding of MR systems. Studies require effective debriefing that helps the participant understand how they were deceived, what elements were deceptive, and how to evaluate potential deceptions. While this SoK synthesizes knowledge from diverse domains, it does not directly involve human-subjects research or development of interactive systems. However, we do provide a framework for exploring cognitive and technological harm of deception attacks in MR.

Open Science

The primary artifact for this SoK is a comprehensive list of articles analyzed during development of the ontology and corresponding MR Deception Analysis Framework. This list includes articles cited in this work as well as others that are not cited. A link to this list can be found at: https://doi.org/10.5281/zenodo.14732979. No other research artifacts, besides diagrams and tables presented in this paper, resulted from this research.

References

[1] Devon Adams, Alseny Bah, Catherine Barwulor, Nureli Musaby, Kadeem Pitkin, and Elissa M. Redmiles. Ethics emerging: the story of privacy and security perceptions in virtual reality. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 427–442, Baltimore, MD, August 2018. USENIX Association.

[2] Eytan Adar, Desney S. Tan, and Jaime Teevan. Benevolent deception in human computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, page 1863–1872, New York, NY, USA, 2013. Association for Computing Machinery.

[3] Abdullah Al Arafat, Zhishan Guo, and Amro Awad. Vr-spy: A side-channel attack on virtual key-logging in vr headsets. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 564–572. IEEE, 2021.

[4] Syed Muhammad Ali, Zeeshan Mahmood, and Dr. Tahir Qadri. 3d view: Designing of a deception from distorted view-dependent images and explaining interaction with virtual world. *Sir Syed University Research Journal of Engineering &amp; Technology*, 7(1):11, Dec. 2018.

[5] Allied Analytics LLP. Mixed reality market to reach $456.8 billion, globally, by 2032 at 67.0% cagr: Allied market research. https://finance.yahoo.com/news/mixed-reality-market-reach-456-140000614.html, 2024.

[6] R.C. Atkinson and R.M. Shiffrin. Human memory: A proposed system and its control processes. volume 2 of *Psychology of Learning and Motivation*, pages 89–195. Academic Press, 1968.

[7] Ankit Attkan and Virender Ranga. Cyber-physical security for iot networks: a comprehensive review on traditional, blockchain and artificial intelligence based key-security. *Complex & Intelligent Systems*, 8(4):3559–3591, 2022.

[8] Alan D. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

[9] Alan D. Baddeley. *Working Memory, Thought, and Action*. Oxford Psychology Series. OUP Oxford, 2007.

[10] Alan D. Baddeley and Graham Hitch. Working memory. *The psychology of learning and motivation: Advances in research and theory*, 8:47–89, 1974.

[11] Andrew Borden. What is information warfare? *Aerospace Power Chronicles*, 1999(11):1–1, 1999.

[12] Evren Bozgeyikli and Lal Lila Bozgeyikli. Evaluating object manipulation interaction techniques in mixed reality: Tangible user interfaces and gesture. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 778–787. IEEE, 2021.

[13] Bo Brinkman. Willing to be fooled: Security and autoamputation in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, pages 89–90, 2012.

[14] James Brown, Jeremy Bailenson, and Jeffrey Hancock. Misinformation in virtual reality. *Journal of Online Trust and Safety*, 1(5), 2023.

[15] Lachlan Brumley, Carlo Kopp, and Kevin B Korb. Cutting through the tangled web: An information-theoretic perspective on information warfare. *Air Power Australia Analyses*, 9(2):1, 2012.

[16] David B. Buller and Judee K. Burgoon. Interpersonal deception theory. *Communication Theory*, 6(3):203–242, 1996.

[17] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. Immersive virtual reality attacks and the human joystick. *IEEE Transactions on Dependable and Secure Computing*, 18(2):550–562, 2021.

[18] Peter Casey, Rebecca Lindsay-Decusati, Ibrahim Baggili, and Frank Breitinger. Inception: Virtual space in memory space in real space – memory forensics of immersive virtual reality with the htc vive. *Digital Investigation*, 29:S13–S21, 07 2019.

[19] Ryszard Cetnarski, Alberto Betella, H. Prins, S. Kouider, and P. Verschure. Subliminal response priming in mixed reality: The ecological validity of a classic paradigm of perception. *PRESENCE: Teleoperators and Virtual Environments*, 23:1–17, 2014.

[20] Yasra Chandio, Noman Bashir, and Fatima M Anwar. Stealthy and practical multi-modal attacks on mixed reality tracking. In *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pages 11–20. IEEE, 2024.

[21] Kaiming Cheng, Arkaprabha Bhattacharya, Michelle Lin, Jaewook Lee, Aroosh Kumar, Jeffery F Tian, Tadayoshi Kohno, and Franziska Roesner. When the user is inside the user interface: An empirical study of ui security properties in augmented reality. In *USENIX Security Symposium*, 2024.

[22] Kaiming Cheng, Jeffery F Tian, Tadayoshi Kohno, and Franziska Roesner. Exploring user reactions and mental models towards perceptual manipulation attacks in mixed reality. In *USENIX Security*, volume 18, 2023.

[23] Edward A. Cranford, Cleotilde Gonzalez, Palvi Aggarwal, Milind Tambe, Sarah Cooney, and Christian Lebiere. Towards a cognitive theory of cyber deception. *Cognitive Science*, 45(7):e13013, 2021.

[24] Norman Dalkey and Olaf Helmer. An experimental application of the delphi method to the use of experts. *Management science*, 9(3):458–467, 1963.

[25] Esrnée Henrieke Anne de Haas and Lik-Hang Lee. Deceiving audio design in augmented environments : A systematic review of audio effects in augmented reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 36–43, 2022.

[26] Dorien De Meyer, Jan Kottner, Hilde Beele, Jochen Schmitt, Toni Lange, Ann Van Hecke, Sofie Verhaeghe, and Dimitri Beeckman. Delphi procedure in core outcome set development: rating scale and consensus criteria determined outcome selection. *Journal of Clinical Epidemiology*, 111:23–31, 2019.

[27] Bella M DePaulo, Deborah A Kashy, Susan E Kirkendol, Melissa M Wyer, and Jennifer A Epstein. Lying in everyday life. *Journal of personality and social psychology*, 70(5):979, 1996.

[28] D. Dionisio, E. Granholm, W. Hillix, and W. F. Perrine. Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38 2:205–11, 2001.

[29] Paul E Downing. Interactions between visual working memory and selective attention. *Psychological science*, 11(6):467–473, 2000.

[30] Norah E. Dunbar, Matthew L. Jensen, Elena Bessarabova, Judee K. Burgoon, Daniel Rex Bernard, Kylie J. Harrison, Katherine M. Kelley, Bradley J. Adame, and Jacqueline M. Eckstein. Empowered by persuasive deception: The effects of power and deception on dominance, credibility, and decision making. *Communication Research*, 41(6):852–876, 2014.

[31] Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1):49–98, 1969.

[32] Randall W Engle. Working memory capacity as executive attention. *Current directions in psychological science*, 11(1):19–23, 2002.

[33] Monique Ernst and Martin P. Paulus. Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry*, 58(8):597–604, 2005.

[34] Habiba Farrukh, Reham Mohamed, Aniket Nare, Antonio Bianchi, and Z Berkay Celik. {LocIn}: Inferring semantic location from spatial maps in mixed reality. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 877–894, 2023.

[35] Gonzalo Munilla Garrido, Vivek Nair, and Dawn Song. Sok: Data privacy in virtual reality. *arXiv preprint arXiv:2301.05940*, 2023.

[36] Joseph P. Gaspar, Redona Methasani, and Maurice E. Schweitzer. Emotional intelligence and deception: A theoretical model and propositions. *Journal of Business Ethics*, 177(3):567–584, May 2022.

[37] Joseph P. Gaspar and Maurice E. Schweitzer. The emotion deception model: A review of deception in negotiation and the role of emotion in deception. *Negotiation and Conflict Management Research*, 6(3):160–179, 2013.

[38] James J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology press, 2014.

[39] Victor A. Gombos. The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs*, 132(3):197–214, 2006.

[40] Andrew M Guess and Benjamin A Lyons. Misinformation, disinformation, and online propaganda. *Social media and democracy: The state of the field, prospects for reform*, 10, 2020.

[41] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. Security and privacy approaches in mixed reality. *ACM Computing Surveys*, 52(6):1–37, oct 2019.

[42] Jassim Happa, Mashhuda Glencross, and Anthony Steed. Cyber security threats and challenges in collaborative mixed-reality. *Frontiers in ICT*, 6, 2019.

[43] Walter Herbranson. *Divided Attention*, pages 1–3. Springer International Publishing, Cham, 2017.

[44] Ray Hyman. The psychology of deception. *Annual review of psychology*, 40(1):133–154, 1989.

[45] William James. *The Principles of Psychology*, volume 1. Henry Holt and Company, New York, 1890.

[46] Joseph Jastrow. The psychology of deception. 1900.

[47] Adrianna C Jenkins, Lusha Zhu, and Ming Hsu. Cognitive neuroscience of honesty and deception: a signaling framework. *Current Opinion in Behavioral Sciences*, 11:130–137, 2016. Computational modeling.

[48] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[49] Polly Kang and Maurice E. Schweitzer. Emotional deception in negotiation. *Organizational Behavior and Human Decision Processes*, 173:104193, 2022.

[50] Carlo Kopp. Information warfare. part 1. a fundamental paradigm of infowar. *Part 1. A fundamental paradigm of infowar*, 4, 2000.

[51] Carlo Kopp. Shannon, hypergames and information warfare. *Journal of Information Warfare*, 2(2):108–118, 2003.

[52] Carlo Kopp, Kevin B. Korb, and Bruce I. Mills. Information-theoretic models of deception: Modelling cooperation and diffusion in populations exposed to "fake news". *PLOS ONE*, 13(11):1–35, 11 2018.

[53] Kiron Lebeck, Tadayoshi Kohno, and Franziska Roesner. How to safely augment reality: Challenges and directions. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*, HotMobile '16, page 45–50, New York, NY, USA, 2016. Association for Computing Machinery.

[54] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. Securing augmented reality output. In *2017 IEEE symposium on security and privacy (SP)*, pages 320–337. IEEE, 2017.

[55] David Leblanc. Dreadful. https://blogs.msdn.microsoft.com/david_leblanc/2007/08/14/dreadful/. Accessed: 2024-08-27.

[56] Tracey Ledoux, Anthony S Nguyen, Christine Bakos-Block, and Patrick Bordnick. Using virtual reality to study food cravings. *Appetite*, 71:396–402, 2013.

[57] Hyunjoo Lee, Jiyeon Lee, Daejun Kim, Suman Jana, Insik Shin, and Sooel Son. AdCube: WebVR ad fraud and practical confinement of Third-Party ads. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 2543–2560. USENIX Association, August 2021.

[58] Timothy R Levine. Truth-default theory (tdt) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4):378–392, 2014.

[59] Timothy R. Levine. Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, 47:101380, 2022.

[60] Ming Li, Xin Chen, Xin Li, Bin Ma, and P.M.B. Vitanyi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, 2004.

[61] Zhen Ling, Zupei Li, Chen Chen, Junzhou Luo, Wei Yu, and Xinwen Fu. I know what you enter on gear vr. In *2019 IEEE Conference on Communications and Network Security (CNS)*, pages 241–249. IEEE, 2019.

[62] Shiqing Luo, Anh Nguyen, Hafsa Farooq, Kun Sun, and Zhisheng Yan. Eavesdropping on controller acoustic emanation for keystroke inference attack in virtual reality. In *The Network and Distributed System Security Symposium (NDSS)*, 2024.

[63] Norman H Mackworth. The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1(1):6–21, 1948.

[64] Steven A McCornack. Information manipulation theory. *Communications Monographs*, 59(1):1–16, 1992.

[65] Richard McPherson, Suman Jana, and Vitaly Shmatikov. No escape from reality: Security and privacy of augmented reality browsers. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, page 743–753, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.

[66] Gregory McWhirter. Behavioural deception and formal models of communication. *The British Journal for the Philosophy of Science*, 67(3):757–780, 2016.

[67] Peter Mell, Karen Scarfone, and Sasha Romanosky. Common vulnerability scoring system. *IEEE Security & Privacy*, 4(6):85–89, 2006.

[68] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329, 12 1994.

[69] Robert W. Mitchell. The psychology of human deception. *Social Research*, 63(3):819–861, 1996.

[70] R.W. Mitchell. *Deception: Perspectives on Human and Nonhuman Deceit*, chapter A Framework for Discussing Deception. SUNY series in animal behavior. State University of New York Press, 1986.

[71] Gillian Murphy, John A Groeger, and Ciara M Greene. Twenty years of load theory—where are we now, and where should we go next? *Psychonomic bulletin & review*, 23:1316–1340, 2016.

[72] Vivek Nair, Gonzalo Munilla Garrido, Dawn Song, and James O'Brien. Exploring the privacy risks of adversarial vr game design. In *Proc. 24th Privacy Enhancing Tech. Symp*, pages 238–256, 2023.

[73] Evangelos Niforatos, Adam Palma, Roman Gluszny, Athanasios Vourvopoulos, and Fotis Liarokapis. Would you do it?: Enacting moral dilemmas in virtual reality for understanding ethical decision-making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.

[74] Niels Christian Nilsson, Tabitha Peck, Gerd Bruder, Eri Hodgson, Stefania Serafin, Mary Whitton, Frank Steinicke, and Evan Suma Rosenberg. 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications*, 38(2):44–56, 2018.

[75] Blessing Odeleye, George Loukas, Ryan Heartfield, and Fotios Spyridonis. Detecting framerate-oriented cyber attacks on user experience in virtual reality. 2021.

[76] Shi Qing Pan, Maria Vega, Alan J Vella, Brian H Archer, and GR Parlett. A mini-delphi approach: An improvement on single round techniques. *Progress in tourism and hospitality research*, 2(1):27–39, 1996.

[77] Michael I Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980.

[78] Michael I Posner and Steven E Petersen. The attention system of the human brain. *Annual review of neuroscience*, 13(1):25–42, 1990.

[79] OU Qiong. A brief introduction to perception. *Studies in literature and language*, 15(4):18–28, 2017.

[80] Irvin Rock. *The Logic of Perception*. MIT Press, Cambridge, 1983.

[81] Franziska Roesner, Tadayoshi Kohno, and David Molnar. Security and privacy for augmented reality systems. *Commun. ACM*, 57(4):88–96, apr 2014.

[82] Somaiieh Rokhsaritalemi, Abolghasem Sadeghi-Niaraki, and Soo-Mi Choi. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, 10:636, 01 2020.

[83] Valerio Santangelo and Emiliano Macaluso. The contribution of working memory to divided attention. *Human Brain Mapping*, 34(1):158–175, 2013.

[84] Susanne Schmidt, Oscar Javier Ariza Nunez, and Frank Steinicke. Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In *Symposium on Spatial User Interaction*, pages 1–10, 2019.

[85] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[86] Cong Shi, Xiangyu Xu, Tianfang Zhang, Payton Walker, Yi Wu, Jian Liu, Nitesh Saxena, Yingying Chen, and Jiadi Yu. Face-mic: inferring live speech and speaker identity via subtle facial dynamics captured by ar/vr motion sensors. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 478–490, 2021.

[87] Herbert A Simon. Information processing models of cognition. *Annual review of psychology*, 30(1):363–396, 1979.

[88] Carter Slocum, Yicheng Zhang, Nael Abu-Ghazaleh, and Jiasi Chen. Going through the motions:{AR/VR} keylogging from user head motions. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 159–174, 2023.

[89] Carter Slocum, Yicheng Zhang, Erfan Shayegani, Pedram Zaree, Nael Abu-Ghazaleh, and Jiasi Chen. That doesn't go there: Attacks on shared state in Multi-User augmented reality applications. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 2761–2778, Philadelphia, PA, August 2024. USENIX Association.

[90] Elizabeth Spelke, William Hirst, and Ulric Neisser. Skills of divided attention. *Cognition*, 4(3):215–230, 1976.

[91] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. Sok: Authentication in augmented and virtual reality. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 267–284. IEEE, 2022.

[92] Courtney Stevens and Daphne Bavelier. The role of selective attention on academic foundations: A cognitive neuroscience perspective. *Developmental cognitive neuroscience*, 2:S30–S48, 2012.

[93] Zihao Su, Kunlin Cai, Reuben Beeler, Lukas Dresel, Allan Garcia, Ilya Grishchenko, Yuan Tian, Christopher Kruegel, and Giovanni Vigna. Remote keylogging attacks in multi-user vr applications. *arXiv preprint arXiv:2405.14036*, 2024.

[94] Robert Trivers. Deceit and self-deception. *Mind the gap: Tracing the origins of human universals*, pages 373–393, 2010.

[95] Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. The dark side of perceptual manipulations in virtual reality. In *CHI Conference on Human Factors in Computing Systems*. ACM, apr 2022.

[96] Yazhou Tu, Zhiqiang Lin, Insup Lee, and Xiali Hei. Injected and delivered: Fabricating implicit control over actuation systems by spoofing inertial sensors. In *27th USENIX security symposium (USENIX Security 18)*, pages 1545–1562, 2018.

[97] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.

[98] Nash Unsworth and Matthew K Robison. Working memory capacity and sustained attention: A cognitive-energetic perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(1):77, 2020.

[99] Martin Vondráček, Ibrahim Baggili, Peter Casey, and Mehdi Mekni. Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses. *Computers & Security*, 127:102923, 2023.

[100] Xian Wang, Lik-Hang Lee, Carlos Bermejo Fernandez, and Pan Hui. The dark side of augmented reality: Exploring manipulative designs in ar. *International Journal of Human-Computer Interaction*, pages 1–16, 2023.

[101] Yingxu Wang. On the cognitive processes of human perception with emotions, motivations, and attitudes. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 1(4):1–13, 2007.

[102] Zhuolin Yang, Cathy Yuanchen Li, Arman Bhalla, Ben Y Zhao, and Haitao Zheng. Inception attacks: Immersive hijacking in virtual reality systems. *arXiv preprint arXiv:2403.05721*, 2024.

[103] Ananya Yarramreddy, Peter Gromkowski, and Ibrahim Baggili. Forensic analysis of immersive virtual reality social applications: a primary account. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 186–196. IEEE, 2018.

[104] Yicheng Zhang, Carter Slocum, Jiasi Chen, and Nael Abu-Ghazaleh. It's all in your head (set): Side-channel attacks on AR/VR systems. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 3979–3996, 2023.

[105] Miron Zuckerman, Bella M. DePaulo, and Robert Rosenthal. Verbal and nonverbal communication of deception. volume 14 of *Advances in Experimental Social Psychology*, pages 1–59. Academic Press, 1981.