University of Southampton

Faculty of Social Sciences

School of Economic, Social and Political Sciences

Scambaiting as Entrepreneurial Digilantism – The Evolution of Scambaiting and its Role in Combating Online Fraud

DOI https://doi.org/10.5258/SOTON/T0073

by

Cosmin Angheluta

Thesis for the degree of Doctor of Philosophy

August 2025

University of Southampton

Abstract

Faculty of Social Sciences

School of Economic, Social and Political Sciences

Doctor of Philosophy

Scambaiting as Entrepreneurial Digilantism – The Evolution of Scambaiting and its Role in Combating Online Fraud

by

Cosmin Angheluta

This research explores the evolution of scambaiting as a form of digital vigilantism, arguing that its modern iteration on platforms like YouTube constitutes a form of "entrepreneurial digilantism." As internet technologies have enabled content monetisation, scambaiting has transitioned from a niche hobby into a structured practice where disrupting fraud is intertwined with creating engaging, revenue-generating content. This study addresses a significant gap in academic literature by examining the performative, economic, and ethical dimensions of this phenomenon.

Employing a constructivist grounded theory approach, the study is analysed through the theoretical lens of symbolic interactionism, with a specific focus on Erving Goffman's dramaturgical analysis. This framework is used to deconstruct the intricate performances staged by both scammers and scambaiters, examining their roles, scripts, and the 'front stage' interactions presented to their dual audiences, the scammer and the online community. The analysis is based on a massive dataset collected from over 5,000 YouTube videos and 6 million comments.

Key findings reveal that scambaiting is a carefully managed performance where scambaiters use deception and role-playing not only to disrupt fraud but also to craft compelling narratives for viewers. The study highlights the crucial role of the audience as a 'performance team,' whose engagement validates the scambaiter’s actions and co-constructs the meaning of the encounter. This research contributes a nuanced understanding of scambaiting as a complex social phenomenon where justice, entertainment, and commerce converge, offering a contemporary perspective on the future of citizen-led responses to cybercrime.

Table of Contents

Research Thesis: Declaration of Authorship

Print name: Cosmin Angheluta

Title of thesis: Scambaiting as Entrepreneurial Digilantism – The Evolution of Scambaiting and its Role in Combating Online Fraud

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: .....Date: 15th of August 2025

Acknowledgements

This research would have not been possible without the support, guidance, and encouragement of many individuals. I would like to express my deepest gratitude to those who have been instrumental in the successful completion of this thesis.

First and foremost, I would like to express my deepest gratitude to my family. The sacrifices made to enable me to pursue my academic interests, and the relentless support, encouragement, and guidance provided have been essential towards my achievements. For this, I am extremely thankful to my parents: Daniel and Liliana, and to my younger brother, Claudiu.

This thesis would have also not been accomplished without the exceptional support provided by my supervisory team. The advice, feedback, encouragement, kindness, and patience provided by Dr Craig Webber and Dr Chris Hamerton have played a key role in the emergence and evolution of this piece of research, with their input having had significant impact on the direction of the work carried out. Additionally, I am also thankful to the Viva panel: Dr Michelle Newberry & Dr Michael McGuire for their invaluable feedback and for the great insights they have brought to this thesis.

I would also like to extend my appreciation and gratitude to the University of Southampton, The ESPS Graduate School Office, Hartley Library, iSolutions, and Student Services Teams for all of the support and assistance offered which facilitated the completion of this PhD. Special thanks also need to be given to my dear friends: Gabriela, Orshi, Artek, Dimi, and Victoria, who have played no small part in keeping me sane throughout this endeavour. I am extremely grateful for their support, encouragement, insights, and, most importantly, friendship.

Finally, I also want to thank Anisa Cox and Lauren Tancock for the excellent support and flexibility offered, which enabled me to keep on top of my PhD while also attending to the responsibilities of my other work within the university. To all those mentioned, and others who have contributed in ways both large and small, my deepest thanks for your unwavering support throughout this journey.

Definitions and Abbreviations

BBC	A British public service broadcaster. Formerly known as the British Broadcasting Corporation. It strives to follow the directive of its founder, Lord Reith, to “inform, educate and entertain”
CGT	Constructivist Grounded Theory – a qualitative research methodology that emphasizes the co-construction of knowledge between the researcher and participants, involving iterative processes of data collection and analysis to develop theory grounded in the data
CRM	Customer Relationship Management – a technology and strategy used by business to manage interactions with current and potential customers, encompassing tools and systems for tracking customer data and enhancing customer service
Digilantism	An adaptation of vigilantism to the digital context, referring to the use of online platforms and tools by individuals or groups to execute acts of vigilantism
KYC	Know Your Customer – a process used by financial institutions and other regulated organizations to verify the identity of their clients and assess potential risks of illegal activities, often involving the collection of personal and financial information
LLM	Large Language Models – advanced artificial intelligence systems trained on large volumes of data to understand and generate human-like text, used for a variety of applications, including natural language processing, text analysis, and data interpretation
RAT	Remote Access Tools – software applications that enable users to access and control another computer remotely over the internet, commonly used by scammers to gain access to victims’ devices in order to extract information or exploit their systems
Scambaiting	The practice of engaging with online scammers, often by pretending to be a potential victim, with the intent to disrupt scammers’ operations, gather information, and expose fraudulent tactics
VoIP	Voice over Internet Protocol – a technology that allows for voice communication and multimedia sessions over the internet, enabling users to make phone calls or engage in video conferencing without traditional phone lines

Chapter 1 Introduction: Scambaiting in the Context of Online Fraud and Digital Vigilantism

1.1 The Rise of Online Fraud and the Emergence of Scambaiting

Over the past years, online scams have escalated into a profound global threat, impacting millions of individuals and organisations worldwide and causing immense economic and emotional harm. The scale of the financial damage is staggering; in the United States alone, the FBI's Internet Crime Complaint Center (IC3) recorded over 12.5 billion in reported losses in 2023 (FBI, 2024). On a global scale, the total cost of cybercrime is projected to reach $10.5 trillion annually by 2025, a figure so significant it is frequently highlighted by organisations like the World Economic Forum as a primary global risk (Cybersecurity Ventures, 2023). Beyond the financial impact, the psychological toll on victims can be severe, with cases of anxiety, depression, and even suicide reported following scam victimisation (Booth, 2024).

In response, a unique form of online vigilantism, known as scambaiting, has emerged. Scambaiters engage with scammers aiming to waste their time, gather information, and potentially prevent future scams by disrupting operations. The practice of scambaiting has existed since the earlier days of the internet but has seen significant expansion in recent years, primarily due to the evolution of internet and communication technologies, as well as online content monetisation models. This has enabled the practice of scambaiting to expand beyond a simple hobbyist endeavour to a more structured and strategic approach, with scambaiters leveraging resources obtained through content monetisation channels to expand and enhance the effectiveness of their operations.

Despite the growth in popularity and the potential contributions that scambaiting brings to the wider landscape of cybersecurity, the academic literature remains sparse and fragmented. While cybersecurity often focuses on technical defences such as firewalls and encryption, scambaiting contributes through social and intelligence-gathering means. It can expose scammers' tactics, gather intelligence on fraud networks, disrupt their operations by wasting their time, and raise public awareness of specific threats, thereby complementing traditional technical security measures. This lack of scholarly attention limits our understanding of the role of scambaiting within the broader contexts of online fraud, cybersecurity, and law enforcement strategies. Furthermore, the effect of content monetisation upon the practice of scambaiting remains largely unexplored. The ethical, legal, and operational consequences – both positive

and negative – of such monetisation practices on the effectiveness and evolution of scambaiting operations remains significantly underexplored.

This thesis aims to bridge the gap in academic research on scambaiting by offering a comprehensive analysis of its practices, evolution, impacts, and implications within the context of modern cybersecurity and online fraud. The hope is to offer a more comprehensive picture of modern scambaiting activities through the lens of vigilantism and digilantism. This will facilitate a better understanding of the emergence of scambaiting and its position in relation to the law enforcement landscape.

This study will employ Erving Goffman's dramaturgical analysis, which uses the metaphor of a theatrical performance to understand social interaction. This framework is exceptionally well-suited to the study of scambaiting, where both scammers and scambaiters engage in elaborate performances, constructing and manipulating realities to achieve their respective goals. By analysing the 'front stage' interactions, the management of impressions, and the various roles played by participants, this research will deconstruct the intricate social dynamics of these encounters. Building on this dramaturgical approach, the broader framework of symbolic interactionism will be used to explore how scambaiters and their audiences construct the meaning of their activities and how these perceptions influence the behaviour of different actors involved, including potential scammers. The main objectives of this research are:

- Expanding the academic literature on scambaiting to develop a more comprehensive picture of its emergence and how it has evolved into its modern form.
- Exploring how content monetisation of scambaiting activities influences the strategies, motivations, and ethical contours of scambaiting activities with consideration for the sustainability and effectiveness of scambaiting operations.
- Understanding how meanings, identities, and roles are constructed and negotiated within the scambaiting community and its audience.
- Investigating the legal and ethical dimensions of scambaiting, particularly in relation to its vigilantism aspects.

1.2 Situating Scambaiting: The Historical Context of Policing and Informal Justice

Recent developments of the monetisation of scambaiting activities having emerged over the past decade, particularly through streaming platforms and social media (such as Twitch and YouTube), but also through crowdfunding platforms (such as Patreon) can situate scambaiting digilantism within the realm of privatised criminal justice. By monetising their efforts,

scambaiters are utilising market mechanisms towards conducting crime prevention and awareness activities. This allows scambaiting operations to not only be self-sufficient in terms of funding, but also to be more directly responsible to their own communities. If individual members of a scambaiting community become dissatisfied with any aspect of the operations carried out, they can simply withdraw their support and funding by unsubscribing or cancelling their recurring contributions.

To better situate scambaiting within the context of modern policing, it is important to take a brief look at the evolution of policing and its current state. Prior to the establishment of formal police forces, informal mechanisms were often the most reliable mechanisms for maintaining order and safety. This would often take the form of community efforts such as watch systems and communal patrols (Johnston, 1996). There was an expectation placed on community members to raise an alarm in response to crimes, and constables, who were elected from among the local community, played a central role in maintaining order and enforcing the law (with the law referring being enshrined more in communal norms rather than formalised legal codes) (Emsley, 2017).

One such practice that was frequently employed consisted of a watch and ward system which required male citizens to guard their towns and cities in order to deter and combat criminal activity. This system was developed as early as the 13th century, with the purpose of maintaining peace and order in local communities, with the watch and ward corresponding to nighttime and daytime surveillance respectively. The Statute of Winchester of 1285 (enacted by King Edward I) reformed the watch and ward system by introducing regularisation, mandating that the watch be kept in every city and town throughout the year (as opposed to the previous ad hoc approach), compulsory service, making it mandatory for all males over the age of 12 to serve in the watch, and by introducing the formalised role of a constable who would be responsible for overseeing the watch and ward system. The reforms introduced by the Statute of Winchester were a significant step in transitioning from an informal feudal system of security to a more centralised and organised system to maintain public order.

Following the introduction of the Statute of Winchester, the responsibility for maintaining order continued to be held by local communities through watch systems and the hue and cry, but these responsibilities were attended to in a more structured and organised fashion. However, the limitations of such an approach began to surface over time with the growth of towns and the increasing complexity of urban life. The need for more effective local governance and law was further accentuated by the Black Death and the consequent shortages of labour and social unrest. The Justice of the Peace Act (1361) came as a response to these needs but also in light of concerns around the reintegration into society of soldiers returning from the war in France

(McBain, 2015). This act serves as another steppingstone on the path of modernizing and formalising policing in the UK by formally establishing Justices of the Peace, a role encompassing both judicial and policing functions. Justices of the Peace were empowered to investigate crimes, arrest and detain suspects, and hold preliminary hearings (Ormrod et al., 2009). These functions expanded further throughout the 17th and 18th centuries, with Justices of the Peace becoming central to administering local justice, overseeing the work of constables and watchmen, and the regulation of the public morals (Harding, 1960).

Another development that occurred in the 18th century was the establishment of the Bow Street Runners in 1749, which is often considered to be a response to the increasing crime rates in the rapidly expanding city of London. This initiative was motivated by a need for more effective mechanisms to combat crime. The Bow Street Runners operated as both investigators and officers, pursuing criminals and recovering stolen property. Most notably, unlike their predecessors, the Bow Street Runners were paid by the government, marking a move towards more professionalised policing (Justices of the Peace were initially voluntary and unpaid positions held by individuals of substantial local standing and wealth, while watchmen were compensated from local parish rates or levies). This group initially established by magistrate Henry Fielding, saw further expansion in size and scope under Sir John Fielding (Henry Fielding's half-brother), who introduced reforms aimed at improving the efficiency of the Runners. Among these reforms, most notably we can observe a rudimentary form of criminal records and the introduction of printed wanted notices (Rawlings, 2012).

The Bow Street Runners were eventually disbanded but their existence served to demonstrate the viability and necessity of a salaried police force dedicated to the proactive prevention and investigation of crime. Their establishment set the groundwork for subsequent reforms and expansion of policing, providing significant influence towards the implementation of the Metropolitan Police.

The transition to a more centralised and organised police force was taken further by the establishment of the Metropolitan Police in London, through the Metropolitan Police Act 1829. This act established under Home Secretary Sir Robert Peel led to the first modern police force in the UK differentiated from armed forces and focused on crime prevention and maintaining public order. This advancement in approaches to policing underscored the necessity of a dedicated, professional body to maintain public order, in light of broader societal changes relating to urbanization and industrialization (Emsley, 2017). This laid the groundwork for modern policing strategies, emphasizing the importance of using a structured and accountable system in order to effectively maintain social order.

The Metropolitan Police Act 1829 was a landmark piece of legislation that established the first professional and centralised police force in London. The early 19th century London was marked by rapid urbanization, increasing crime rates, and social unrest which highlighted the inadequacies of existing policing systems and the need for a more effective and organised approach to law enforcement. Central to the introduction of the Metropolitan Police Act was Sir Robert Peel (who was the Home Secretary at the time). In drafting the act, Sir Robert Peel showed consideration for prevention of crime over punitive measures, the importance of police-citizen relations, and the efficiency of a uniformed police force. The introduction of the Metropolitan Police Service replaced existing local watch systems and marked the first instance of officers (known as “Peelers” or “Bobbies”) being selected based on merit, being awarded salaries, and being subject to training to equip them for their roles.

One of the focus points of the Metropolitan Police was the emphasis on preventing crime rather than a reactionary approach of responding and punishing it. Uniformed officers were tasked with patrolling the streets, serving as a proactive deterrent to crime. The perceived success of the Metropolitan Police led to the establishment of similar forces in other cities and regions, building towards the gradual establishment of a national policing system. This represented a significant shift from community-based approaches towards maintaining peace and order to a state-run professional service funded and dedicated to these endeavours (Emsley, 2017).

The expansion of professional policing across the country was encouraged by the County Police Act of 1839 which came about in recognition of the need to extend professional policing to rural areas and smaller town where policing was still managed by parish constables who were often part-time and unpaid. The County Police Act empowered Justices of the Peace to form professional police forces, funded by local taxes, but it did not make such police forces mandatory, leaving it to the discretion of the local taxpayers. Subsequent acts such as the Rural Constabulary Act of 1840 and the County and Borough Police Act of 1856 made the establishment of police forces mandatory throughout England and Wales and subjected them to periodic inspections to ensure certain standards of efficiency. The County and Borough Police Act also introduced government funding towards the costs of local police forces, contingent on the passing of standards set by the Home Office inspectors.

The acceleration of technological advancement that began after the Industrial Revolution provided further opportunities, as well as challenges, for the evolution of policing operations. Improvements in communication technologies, such as the advent of the telegraph, radio communications, and, later, the telephone, enhanced communication between police forces and across jurisdiction allowing for quicker response times and facilitating better sharing of information about crimes and suspects across wider geographical areas (Briggs and Burke,

2009). Historically, police agencies began integrating telegraph alarm systems into municipal infrastructure to synchronize operations during periods of urban expansion (Tarr, 1992), and over the decades, these systems evolved to include radio communication, reshaping organizational practices and real-time operational responsiveness (Soullière, c. 1999).

In the modern era, the deployment of data-driven systems, including GIS, mobile communication devices, and networked information platforms, has further streamlined inter-jurisdictional coordination, dispatch, and suspect tracking (Ugwudike, 2024). Numerous forensic advances have also been seen throughout the 20th century, with some examples including the adoption of fingerprinting, and the development of DNA analysis. Numerous forensic advances unfolded during the 20th century. The widespread institutional adoption of fingerprinting, established as reliable and effective from around 1901 (Polson, 1951), has continued evolving into modern digital and automated identification systems (Fingerprint Science review, 2023; Oxford entry on colonial fingerprinting, 2024). Later, the revolutionary introduction of DNA analysis transformed forensic investigations: foundational reviews document the three-decade arc of method refinement (Roewer, 2013), technical expansions into STR and SNP markers (Bukyya et al., 2021), and ongoing scientific and ethical debates surrounding DNA implementation (Oosthuizen, 2022).

The latter half of the 20th century also continued to see an expansion of the formal policing model with a shift towards greater public accountability and the concept of community policing. Significant reforms were introduced through The Police and Criminal Evidence Act of 1984 (PACE) which aimed to ensure fair treatment and to protect the rights of individuals during formal police activities. Efforts were also put towards building trust between the police and the communities they serve, emphasizing partnership and problem solving towards the maintaining of peace and order. Furthermore, police forces underwent significant expansion and professionalisation, incorporating advancements in technology and forensic science which enabled and improved their capabilities to detect and prevent crime (Brodeur, 2010).

In the late 20th and early 21st century, a rise of neoliberal approaches to policing can be observed, moving towards a more market-centred approach to governing social and economic life, and a reduced level of state intervention. The role of private entities has also been seeing an increase in areas traditionally managed by the state, partly due to a need to keep up with technological advances, and partly due to the increasing focus on efficiency and effectiveness.

Privatisation can be observed in a number of policing functions with entities such as private security companies and private forensic laboratories fulfilling responsibilities previously held by the police, which are now being outsourced (Montgomery & Griffiths, 2016; Finn, 2021; Forst, 2000). The development of partnerships between public police forces and private entities

became increasingly necessary reflecting the neoliberal ethos of collaboration between the state and the market. Some of the elements that contributed to this movement include an increased focus on performance metrics for policing, attempting to improve efficiency, accountability, and to implement different measures of success, shifting into a more results-based form of governance. In line with this, pushes have also been made to make policing more cost-effective with resources being allocated based on analyses of crime trends and risks, seeking to optimise the use of the limited available resources. Neoliberal approaches to policing have also placed greater emphasis on risk management strategies, focusing on identifying and mitigating risks before they manifest as crimes, most recently employing the use of technology and data analytics to predict crime trends. (Loader & Walker, 2007)

This indicates a shift from the principles initially envisioned by Sir Robert Peel which were grounded in fostering community trust, aiming to integrate community officers into the community rather than positioning them as an external force. Additionally, the marketization and privatisation of policing is also moving away from policing as a purely public service, with an increasing focus being placed on market solutions rather than public service. Furthermore, while the importance of crime prevention is apparent both in Sir Robert Peel's initial vision and in the more recent neoliberal approaches, the former is more engrained in community-oriented practices, while the latter is taking a more targeted approach based on risk assessments and economic rationales.

1.3 Contemporary Challenges: Cybercrime and the Limits of Formal Policing

The contemporary landscape of formal policing is faced with a multitude of challenges, both to do with efficiency due to resource allocation, but also with public perception. Budget constraints and resource allocation have been persistent challenges for the UK police, affecting their ability to effectively respond to all types of crime. Austerity measures on British policing have resulted in a need to be more selective on the use of resources available and the need to be strategic in the prioritisation of crimes (Loader, 2013). McGuire (2016) also notes on how the increasingly rapid technological advancements, alongside the transnational nature of cybercrime create increasingly difficult challenges for law enforcement, especially considering resource constraints stemming from budget limitations and training gaps.

Technological advancements and the emergence of cybercrime have also contributed to the plethora of challenges faced by modern policing. The rapid proliferation and expansion of the internet has greatly transformed criminal activities, enabling specific offences (such as fraud) to occur at greater scale and across jurisdictions, making it difficult for local or national law

enforcement to intervene effectively (Wall, 2024). A perceived surge of cyber-induced problems has sparked an accompanying wave of social policy activity, with a key example being the UK's Online Safety Act, which aims to regulate online platforms and protect users from harmful content. This development highlights a critical debate in criminology; many online frauds, like advance-fee fraud or phishing schemes, are digital versions of traditional white-collar crimes such as fraud and embezzlement. The debate centres on whether the 'cyber' element constitutes a new category of crime or simply a new tool for old offences, a question highly relevant to understanding corporate and financial crime in the digital age.

However technological advancements also present a challenge through the manner in which they might be adopted by law enforcement. Increasing adoption of technologies such as predictive policing algorithms and facial recognition, while carrying the potential of improving efficiency, serving to counterbalance issues around adequate resourcing, also present a risk of overreach and misuse, creating a risk of distancing policing from more human elements such as community engagement and discretionary judgement (McGuire, 2021). The increasing demand placed by rapidly evolving cybercrime on law enforcement adds pressure towards the adoption of increasingly sophisticated technological solutions, which, out of a desire for rapid results, might bypass necessary ethical and practical considerations. In this technological arms race, fraudsters also employ their own tools, such as the use of VoIP for anonymous calling (Metropolitan Police Operation Elaborate, 2022; Steinmetz & Holt, 2022), email spoofing to impersonate legitimate entities (Varshney, Misra & Atrey, 2016), and social engineering toolkits to carry out their schemes (Song, Kim & Gkelias, 2014).

Looking specifically at the issues of fraud and scams, the UK has seen increased levels of privatisation and collaboration with the private sector towards tackling such risks. Most notably, a significant part of the efforts to combat fraudulent activities is currently sitting within the financial sector with banking companies. Examples of such efforts consist of bank-led initiatives to introduce sophisticated systems for the detection and prevention of fraudulent transactions (including real-time fraud detection algorithms and customer alert systems), due diligence and customer verification training for banking staff members to enable them to identify potential instances of customers being defrauded and to intervene, as well as banking protocols allowing bank staff to flag suspicious activities to local police via rapid response system. Furthermore, banks and financial companies are increasingly held to higher standards of customer care and due diligence to prevent fraud, some of these standards translating into industry-wide regulatory requirements. While these institutional efforts form a critical line of defense, scammers have actively developed social engineering tactics to circumvent them.

This approach of tackling fraud and scams through a mix of public-private partnerships, banking regulations, and direct interventions by financial institutions has seen both successes and challenges. UK Finance's Annual Fraud Report (2023) details that over £1.2 billion was stolen through fraud in 2022, which is a reduction of approximately 8% compared to 2021. Some of this reduction is attributed to protections such as SCA (Strong Customer Authentication) and Confirmation of Payee. Furthermore, public awareness campaigns such as the "Take Five to Stop Fraud" campaign launched by UK Finance contribute to wider public awareness about the risks of fraud and encourages individuals to pause and think before completing financial transactions (UK Finance, 2025; Jensen et al., 2024).

However, challenges continue to exist. The rapidly evolving nature of scams continues to pose challenges with fraudsters learning and adapting to the latest countermeasures implemented. One relatively common strategy that scambaiters have come across consists of scammers attempting to create distrust towards banking staff or law enforcement officials in their victims by advising that such officials are involved in various nefarious activities and cannot be trusted. This claim is supported by reports from anti-fraud organisations which note this as a common social engineering tactic (e.g., Action Fraud, 2023), and supported in scholarly studies showing scammers impersonating authority figures, including police and bank staff (Liu et al., 2023; Wood et al., 2023), exploiting cognitive biases that make victims susceptible to such manipulations (Montañez et al., 2020), and weaponizing emotional trust dynamics (Anderson, 2024). This is intended to reduce the efficiency of mechanisms puts in place for banking staff or law enforcement officials to identify fraudulent activity by preventing the victim from disclosing information that would raise red flags.

In the context of the evolution of UK policing that has been briefly explored above, scambaiting presents an interesting phenomenon. As will become more apparent later on, some scambaiting activities are akin to early watch and ward practices, with scambaiters occasionally monitoring systems within scamming operations and intervening to disrupt scammers' activities, to prevent potential victims from being taken advantage of, and to gather intelligence. However, at the same time, some elements of scambaiting operations are aligned with neoliberal approaches with scambaiters efficiently utilising market mechanisms and solutions to enable and expand their activities, both in terms of disrupting scamming operations, but also in terms of promoting and educating the public on how to protect themselves.

1.4 Traditional and Digital Vigilantism

Vigilantism as a phenomenon has been a part of human society since its early days, arising in circumstances where formal mechanisms of law enforcement were seen as inadequate, or were entirely absent. At its core, vigilantism involves individuals or groups taking the law into their own hands, without formal legal authority and without official sanction, typically in response to perceived threats or injustices. Prior to the establishment of state-sanctioned, professionalised institutions like public police forces (e.g., the Metropolitan Police in London in 1829), justice was often administered through less formal, community-based or private means, and vigilantism can be seen as an extension of these traditions in contexts where state mechanisms are perceived as failing.

Les Johnston (1996) has posited that there are six criteria that an activity needs to fulfil in order to be considered an act of vigilantism: (1) A certain level of planning, premeditation, or organization needs to occur from the individual(s) responsible for the act; (2) The act must be taken by private agents, unaffiliated with law enforcement (although an argument can be made about those affiliated with law enforcement acting outside of this affiliation); (3) The act must be autonomous and free from state support, it should not be sanctioned by the government or by agencies in the criminal justice system; (4) The act must encompass the use of, or threatened use of force (Smallridge et al. (2016) argue that this criteria can be broadened to also include causation of harm in order to fit in with the more modern digilantism); (5) The vigilante activity needs to occur in response to a criminal or socially deviant event that transgresses institutionalized norms. This transgression can be either real or perceived; (6) The goal of the vigilante act must be to control crime or social deviant actions, specifically done by offering assurances of personal and collective security to people.

The roots of vigilantism can be traced back to pre-modern societies where formal state authority was minimal and community enforcement was the main channel of protecting and maintaining norms and laws. For example, during medieval times, “hue and cry” (which has been discussed in the previous chapter) was prevalent across Europe with there being an expectation upon members of a community to make noise and rally their neighbours in pursuit of criminals (Johnston, 1996). Other examples of vigilantism have also been observed in the US, particularly during the 19th century on the frontier and in rapidly developing territories where formal law enforcement was inadequate or absent altogether. A prime example is the California Gold Rush where miners organised committees to combat rampant crime that legal authorities were unable to keep under control (Brown, 1975). These committees often conducted trials and executed punishments ranging from banishment to lynching, thus filling the gap created by inadequate governmental structures. Another example can be found in Peru where “rondas

campesinas" (peasant patrols) emerged in the 1970s as a response to cattle theft, and later evolved to combat the insurgent threat of the Shinning Path guerilla group. In contrast to the type of vigilantism that emerged during the American frontier period, which was more transient and ad-hoc, the community patrols that emerged in Peru were more deeply rooted in local customs and social structures (Starn, 1999).

Traditional vigilantism is often characterized by a few common features among which list a strong sense of community justice, immediacy of action, and direct methods of punishment. The justification for such actions usually hinges on the community's perception of the ineffectiveness of legal institutions, corruption, or the urgency of needing to restore social order. Historically, vigilantism has been seen as both a symptom of an insufficient justice system, and a grassroots method for enforcing social norms (Abrahams, 1987). As societies evolved and formal law enforcement systems became more robust, vigilantism has either declined or transformed. However, the fundamental impulses driving vigilantism (dissatisfaction with formal mechanisms of justice) continue to manifest in modern contexts, taking new forms employing digital tools and platforms, reflecting broader changes in technology, communication, and social organization.

The rise of digital platforms such as social media, online forums, and other internet-based communication tools constitute a crucial element that marks the transition from traditional vigilantism to digilantism (Trottier, 2017). These platforms are great enablers of social interactions, facilitating rapid information dissemination, allowing individuals to reach large audiences almost instantaneously, a capability that was previously limited to traditional media outlets such as television, radio, and printed press. Furthermore, digital technologies provide users with varying levels of anonymity and global accessibility that fundamentally alters the dynamics of vigilantism (Wall, 2007). The availability of such anonymity can empower individuals to act without being as concerned for direct consequences, which could potentially lead to irresponsible and harmful actions undertaken in the absence of the accountability usually enforced in physical settings.

Neil Postman (1992) posits that human experiences, perceptions, and society as a whole are fundamentally shaped by media and communication technologies. The advent of the internet and digital communication tools marks a significant ecological change, altering how individuals and communities form, function, and how they engage in social control and justice-related activities. In this context, digital tools go beyond being mere tools and can be viewed as environments that reshape social dynamics, allowing for rapid dissemination of information and coordination of action beyond the limitations of geographical constraints. Furthermore, McCarthy & Zald (1977)'s concept of Resource Mobilisation can be utilised to gain insights into

how digital tools have transformed the organizational structures and strategies available for vigilante activities. Resource Mobilisation Theory emphasizes the importance of resources in the success of social movements rather than the grievances that might motivate participants. The internet is able to provide unprecedented resources for mobilisation and coordination, lowering the barriers to entry for participation in vigilante acts and further enabling spontaneous and decentralised movements that can operate in the absence of formal organizational hierarchies.

The availability of anonymity and pseudonymity also play a significant role in affording individuals access to participate in vigilante activities. Goffman's Theory of Identity (Goffman, 1959) suggests that anonymity detaches actions from offline identities, potentially emboldening individuals to engage in behaviours that they might avoid under the surveillance of a known community. The usual checks and balances that are generally available can be circumvented, allowing individuals to bypass consequences they might otherwise incur were their identities to be known.

Furthermore, changes in media technologies can themselves lead to shifts in societal norms and ethics (Christians et al. 2009). As digital platforms continue to evolve and redefine the boundaries of privacy and publicness, they also challenge traditional legal and ethical conceptions of justice and punishment. The public's engagement in digilantism is often reflective of these shifting norms regarding privacy, justice, and participation in the digital age.

Digilantism has emerged in recent years as a significant phenomenon where individuals and groups leverage online platforms to enact justice or punishment outside of formal legal systems (Doležal, 2024). The concept of digilantism encompasses a wide range of online actions, including targeted surveillance, dissuasion, or punishment, often in response to perceived offenses (Loveluck, 2019). Attempts have been made at typologising digilantism, distinguishing between flagging, investigating, hounding, and organised leaking, each representing a different mode of online self-justice (Loveluck, 2016).

Digilantism has also been seen by some as a form of consumer activism, with individuals using their online presence to express dissatisfaction with organizational service performance. The manifestations of this are varied, taking the form of shaming, calling for action, or promoting certain behaviours in response to a breach in consumer performance expectations (Legocki, et al. 2020). On the other hand, others have noted the role that digilantism plays in wider socio-politico-economic issues. Tanner et al. (2020) notes how digilante activities are being carried out in Quebec to enforce views on immigration, national identity, and cultural values, particularly through social media channels. One consequence of these practices is an increased polarisation between social groups. Similarly, Huang (2021) discusses how

digilantism has been utilised in China to tackle issues relating to unethical behaviour and violations of norms, such as disrespect, incivility, animal brutality, and others. The channels through which digilantism has been perpetrated has also been observed to evolve alongside technological and socio-political developments, with digilante actors increasingly leveraging platforms such as Sina Weibo and WeChat. Gabdulhakov (2018) also touches upon manifestations of digilantism within Russia, where there is a level of institutionalization to digilante activities, with parallels drawn between modern online exposure tactics and historical surveillance methods (like naming-and-shaming) prevalent during the communist period.

Over the past couple of decades, a number of digilantism activities have been observed, attracting significant media attention, and demonstrating the resourceful utilisation of online tools and platforms towards tackling crime and deviance. One of the most notable examples relates to the Boston Marathon Bombing from 2013. In the aftermath of the incident, users of various social media platforms (including Reddit) took it upon themselves to identify the perpetrators putting together various pieces of information that they were able to gather through various means. This crowd-sourced detective work has unfortunately not been as efficient as the users involved might have hoped, leading to the misidentification of several individuals which resulted in severe consequences for them and for their families (Madrigal, 2013).

Another example of digilantism can be found in Operation Payback, which was a series of coordinated attacks triggered by members of Anonymous targeting major credit card companies and payment gateways (including PayPal), which has stopped processing donations to WikiLeaks (Sauter, 2014). Initially this series of attacks were undertaken to support free information distribution on the internet, but it evolved into a broader movement against intellectual property law, leading into widespread debate around internet governance, freedom, and the ethics of digilantism.

It is apparent that there are numerous parallels between digilantism and traditional vigilantism. Motivations for stepping outside formal mechanisms of law enforcement appear to be similar in that they involve an element of maintaining moral standards and responding to perceived inadequacies. However, traditional vigilantism appears to have been primarily concerned with and rooted in immediate community concerns, such as protecting property and maintaining peace and order within corresponding social groups, whereas digilantism often extends its goals to issues that are more global in nature, transcending local communities (at least insofar as local communities are identified geographically rather than digitally). The scale of digilantism as opposed to traditional vigilantism is therefore much larger, with greater numbers of actors being able to participate and with injustices being targeted outside of geographical limitations.

Differences are also evident in the methods employed with traditional vigilantism involving more physical actions such as capturing, punishing, or deterring criminals through physical presence and, sometimes, through the use of violence (Brown, 1975). Digilantism on the other hand moves towards the utilisation of online tools and platforms to expose, harass, or to mobilise public sentiment against individuals or organizations. Some common tactics can include doxing, shaming, and cyber-harassment, which do not require physical presence, but which can have widespread psychological impact (Douglas, 2016). Furthermore, the digital emphasis extends to the targets of digilantism as well with network infrastructure and resources being put in the line of fire in order to cause disruption or to send a message.

Digilantism also differs from traditional vigilantism in the impacts that result from such activities. With traditional vigilantism, the impacts are often immediate and localized, affecting primarily the direct community in which the vigilante actions occur. Such impacts could consist of increased security and order within a community, or they could be detrimental, causing fear and chaos depending on the nature of the vigilante acts (Abrahams, 2000). On the other hand, the impact of digilantism is often broader and farther-reaching with the potential of reaching a global scale, affecting individuals' lives and reputations worldwide. Digital footprints are also potentially indelible, leading to long-term or permanent damage to individuals' or entities' reputations (Jane, 2016). It is worth noting however, that traditional vigilantism also has the potential to be impactful at a larger scale, especially if we consider some of the examples of vigilante strategies that informed formal law enforcement approaches discussed in the previous chapter.

Traditional vigilantism is also more susceptible to significant legal consequences due to the more direct level of involvement of vigilante actors in their unlawful activities. Such activities are also more easily situated within clear legal boundaries making law enforcement scrutiny more available. Digilantism on the other hand is often more complex and ambiguous, transcending local jurisdictions and legal frameworks and posing a greater challenge for formal law enforcement to tackle. This translates into a more inconsistent and often ineffective level of regulating and combating digilantism (Wall, 2007).

In transitioning from its traditional form, digilantism has retained some of the key motivations and circumstances that can influence individuals to engage in such activities. However, the methods, scope, and implications have seen significant evolutions that reflect broader social, technological, and cultural shifts, which continue to evolve and adapt alongside advances in technology and shifts in social norms across the internet.

1.5 A Comparative Analysis of Vigilantism and Informal Policing

To fully grasp the modern manifestation of digilantism discussed previously, it is useful to first trace its historical roots in earlier forms of informal policing. Vigilantism itself has a long history stretching across various cultures and epochs. Gaining a better understanding of these early forms of vigilantism would be appropriate in building a foundation for the modern manifestation of such practices. Arguably, some of the earliest forms of vigilantism predate the very term and are embedded within social and judicial practices of ancient societies. However, they do share the feature of community members taking justice in their own hands in the absence (or perceived inadequacies) of formal legal authority.

Forms of vigilantism have been noted in tribal societies, where community justice is enacted without a formal legal system, with actions ranging from retributive justice practices to communal enforcement of social norms. (Boehm, 1987). Some tribal societies even operate under a system of blood revenge or feuding, where families or clans take it upon themselves to avenge wrongs committed against their members. This form of justice is deeply rooted in the social fabric of such communities but also in their honour codes (Black-Michaud, 1980).

Instances of vigilantism are also present in Ancient Greece, where a reliance on personal revenge and family feuds to resolve conflicts is often observed prior to the development of formal legal codes (Gagarin, 2015). Avenging a wrong done to a family member was not only a right, but it could even be perceived as a duty.

Some ancient legal codes even included principles allowing victims, or their families, to play a direct role in the execution of justice (lex talionis – the law of retaliation) which, while arguably not fully qualifying as vigilantism (since this was formally included in some form of legislation being applied), it does resemble it in allowing some of the responsibility for carrying out justice to be placed on individuals affected, rather than entirely on a formal judicial mechanism.

More recent examples of vigilantism can be found in the United States between the late 18th and early 20th centuries. In the context of the expansion westward during the 19th century, cast areas with limited law enforcement came to be. This vacuum led to the emergence of vigilante groups, one of the most famous examples consisting of the San Francisco Committee of Vigilance in the 1850s. This group emerged in response to rampant crime and corruption and groups such as this often justified their actions as necessary towards the preservation of order and justice in the absence of effective governmental solutions (Brown, 1975).

Another much more famous vigilante group, originally formed by the ex-Confederate soldiers in the southern US, is the Ku Klux Klan (KKK) whose activities targeted African Americans, Jewish

people, and other minorities, with the purpose of protecting and preserving the interests of white Americans. (Parsons, 2015)

Returning to the UK, another example of vigilantism can be found in the late 19th century, during the industrial revolution, where trade unions would sometimes engage in activities that could be classed as vigilantism, such as enforcing strikes or boycotting employers through picketing. Such actions were taken towards protecting workers' rights and interests in the absence of sufficient legal protections. (Prothero, 2013).

In more recent years, there have been instances of vigilante groups in the UK targeting suspected paedophiles and criminals, often leveraging the internet to identify and expose such individuals or to arrange meetings with the intention of detaining them until law enforcement would arrive to take over. Such groups, usually labelled "paedophile hunters" have gained significant media attention for their efforts to trap and expose individuals seeking to engage in sexual activities with minors (Frampton, 2022).

Instances of vigilantism have been present throughout history in different forms, and the classification of an activity as vigilantism is highly dependent on the social and judicial applicable. While some of the activities of trade unions in the UK in the late 19th century could be categorised as vigilantism in the appropriate temporal context, in the modern day, such activities could be carried out within the protection of legal frameworks that recognise the rights of workers to organise, strike, and engage in collective bargaining. Furthermore, the contribution of early labour movements and their vigilante activities towards the development of legal frameworks that recognise and protect the right to strike is significant (Thompson, 2016)

Furthermore, parallels can be drawn between some of the examples of vigilantism we have visited and the early mechanisms of informal policing that have been adopted by communities towards preserving peace and combating crime. The example of the San Francisco Committee of Vigilance bears similarities to early emergence of watch and ward strategies, where the necessity for policing solutions in the absence of formal options has given rise to vigilante solutions to address the needs identified.

At the same time, vigilante activities have also been observed to be more detrimental than effective, with a good example being provided by the KKK. While the KKK can be classified as a vigilante group, their operations were focused on combatting perceived criminal and socially deviant activities that were rooted in racial prejudice rather than in response to specific actions taken by those in their crosshairs. Such racially prejudiced targeting is more generally recognised in the modern day as socially unacceptable, and judicial progress has also been made towards ensuring fairer provisions for individuals despite differences in race, gender, or

other protected characteristics. However, advances have not been made uniformly across the world, and the activities of the KKK provide an excellent example of vigilantism gone wrong.

The attention attracted by paedophile hunter groups can be attributed, at least in part, due to the complex ethical and legal implications of their approaches. The activities of such groups have on been successful in gathering evidence against individuals who were actively seeking to exploit minors and their actions have led to arrests and convictions, filling the gaps where traditional law enforcement might lack the resources or capabilities to intervene effectively (Lanning et al., 2014). However, there have also been instances where individuals targeted by paedophile hunters were later found to be innocent and the accusations that were brought upon them had devastating effects including reputational damage, mental health struggles, and, in more severe cases, suicide.

The activities of paedophile hunters have also seen criticism due to the methods employed having the potential of compromising ongoing police investigations. Their actions could potentially alert suspects, providing them with the opportunity to destroy evidence or go into hiding. Furthermore, the methods used can also create legal complications to prosecuting suspects relating to provisions around entrapment and admissibility of evidence (Frampton, 2022).

The duality of paedophile hunters' activities raises significant ethical and legal questions. While their intentions might be to protect children and their wider communities, the lack of legal authority, oversight, and adherence to due process poses risks to the integrity of the justice system and the rights of individuals. The balance between community safety and legal and ethical standards of justice is often central to the debate around paedophile hunter activities.

Of all the examples of vigilantism that we have looked at, paedophile hunters are the only one that could fall in the category of digilantism, due to the extensive employment of digital tools and resources that is observed in their activities. Digilantism has become increasingly common with the advent of the internet and social media, which enable a wide range of activities and empower individuals and groups a wider reach and a wider range of resources in conducting vigilante activities. Focusing specifically on digilantism, there are a number of other examples outside of paedophile hunters, such as doxing, hacktivism, and social media shaming.

The evolution of digilantism can be traced alongside the growth of the internet and digital communication technologies. It has also continued to evolve alongside technological advances with more recent forms making extensive use of social media and mobile internet technologies. Identifying the earliest forms of digilantism can be challenging but it could reasonably be expected that it can be traced back as early as the emergence of bulletin board systems where

communities formed around shared interests. The anonymous nature of these platforms allowed users to share information and, occasionally, coordinate actions against individuals or entities perceived as threats or violators of community norms. One example of this would consist of how bulletin board systems would sometimes take it upon themselves to police their spaces against abusive users, which would sometimes translate into collective efforts to remove the access of such users and sharing their information with other operators to prevent them from causing harm elsewhere or using technical means to block their access. Bulletin board systems have also been used for the purposes of activism and information sharing, educating users to bypass government censorship in various countries to counteract perceived injustices under the form of suppression of information. One example of this consists of the use of FidoNet in the late 1980s and early 1990s by users in the Soviet Union and Eastern European countries to bypass state-controlled media and censorship (Milligan, 2023).

As the internet became more accessible, groups began to use hacking as a form of political activism or protest, targeting organizations, governments, and individuals to draw attention to various causes. The 1990s and early 2000s saw the emergence of hacktivism with notable hacktivist groups such as Anonymous or Cult of the Dead Cow beginning to gain notoriety (Jordan et al., 2004).

Social media has fundamentally altered the dynamics of digilantism by enabling swift mobilisation, information sharing, and collective action in ways previously unattainable through traditional offline means. Scholars highlight how digitally mediated vigilantism transcends offline boundaries, using platforms to enact parallel forms of justice (Favarel-Garrigues, 2020; Trottier, 2017). Algorithms amplify visibility and prompt widespread mobilisation (Allen, 2020), while social media's connective power supports the rapid formation of collective identities and action across geographies. The formation of groups focused on exposing wrongdoings was also made more accessible by social media with users becoming more easily able to engage in such activities and access such groups. Social media has also enabled new forms of digilantism to become more widespread such as doxing and online shaming due to platforms such as Facebook encouraging users to utilise identifiable information in the creation of their accounts in order to utilise the features and functionality offered by the platforms. (Trottier, 2017).

Most recently, the technological advances around mobile devices, mobile internet, and real-time video streaming, have enabled digilantism to expand even further with users gaining the ability to live broadcast protests, crimes, and even acts of vigilantism with significant ease (Altheide, 2018).

Similarly to the evolution of digilantism itself, scambaiting has also been evolving alongside technological advances. One of the earliest and most well-documented forms of scambaiting

relates to the so-called “Nigerian prince” scam emails which involve sending mass emails claiming the recipient can claim a significant sum of money by assisting in transferring funds out of a country, for which the scammer promises a large reward but requires upfront payment of fees or personal information. Internet users began to recognise these scams and, instead of ignoring them, some chose to engage with the scammers in order to waste their time, resources, and ideally preventing them from targeting more vulnerable individuals. These scambaiting attempts were often shared on forums and early websites dedicated to the practice, where individuals would post email exchanges or details of their encounter with scammers for both educational and entertainment purposes. One of the pioneering and most well-known early scambaiting communities was the website 419eater.com, which became a central hub for documenting interactions, sharing strategies, and celebrating successes. Scambaiting has continued to evolve with scambaiters making use of the most recent technological advancements available (and appropriate) in order to tackle and educate on modern types of online scams. Scambaiters have leveraged cutting-edge technologies, notably AI-powered chatbots, LLM-driven scam response systems, and custom-built tools, to actively engage scammers, disrupt scam operations, and raise public awareness. Notable advances include ChatGPT-based automatic scam-baiters that significantly increase engagement with scammers (Bajaj & Edwards, 2023), ScamGPT-J which simulates scammer language to help users detect fraud (Tan, See & Kok, 2024), and “Apate” AI bots that pose as potential victims to occupy scammers and collect intelligence (Kaafar et al., 2024).

In the wider context of digilantism, we can see some similarities between scambaiting and other digilante activities that we explored earlier. Due to the challenges posed by the modern digital environment, online scams are challenging to tackle through official policing efforts, leaving a gap for scambaiting to emerge as a digilante activity.

1.6 The Role of Netizens in Digilantism

Central to scambaiting, and to broader forms of digilantism such as the ones mentioned above, is the concept of netizen, which is a portmanteau of “internet” and “citizen”. The term “netizen” was coined to describe individuals actively engaged in online communities, particularly in the realms of communication and information dissemination (Rheingold, 1993). With the evolution of digital platforms and the technologies, the role of the netizen has expanded into various forms of cyber participation, including digilantism (Trottier, 2017). Netizens are not just users of the internet, but rather participants who contribute to the digital arena, engaging in online social actions, disseminating information, and shaping their communities.

In a study looking at netilantism, Chang (2020) found that there are a number of elements that can contribute to a positive perception of digilante activities. Internet users who are willing to engage in netilantism and contribute to Human Flesh Searching (crowd sourced distributed researching activities conducted using online platforms such as blogs and forums) are those who have less confidence in the effectiveness of criminal justice systems. Additionally, positive sentiments towards social justice were also found to be a positive predictor of inclinations towards digilantism.

An important distinction needs to be made between active and passive netizens within the context of scambaiting. Active netizens are more directly involved in interactions with scammers, often employing various strategies (ranging from simple engagement tactics to more complex deceptive approaches) with the goal of extracting information relating to the pertinent scamming operation, or to disrupt scammers' activities and waste the limited resources they have available. Passive netizens on the other hand provide a more indirect contribution to the scambaiting ecosystem, engaging primarily in disseminating information, educating others on combatting scams, and supporting active netizens by providing tools, platforms, information, and more recently, even financial support to enable scambaiting operations.

On the active netizens front, the motivations behind such proactive engagement often reflect a mix of altruism and personal sense of justice. Active netizens frequently engage in activities that seek to protect the vulnerable, but they might also be driven by "the thrill of the hunt" and the intellectual challenges that come with disrupting scamming operations. The scambaiting communities emerging around scambaiters contribute to the reinforcement of such motivations where the celebration of successful scambaiting attempts within the relevant online communities can strengthen the active netizens' identities as combatants of online fraud (Sorell, 2019).

Passive netizens on the other hand might see stronger motivations in a sense of community, solidarity, and commitment to raising awareness that emerges within scambaiting communities. Despite not engaging directly with scammers, their role is quintessential to the maintenance and sustainability of modern scambaiting approaches by providing the information and moral support necessary for active netizens to continue their operations, but also, more recently, by providing financial support to enable scambaiting activities (Denning, 2010). Furthermore, as we will see later in more detail, within the realms of scambaiting, active netizens have been observed to attempt to provide channels of becoming more involved to passive netizens.

The dynamic that emerges between active and passive netizens is akin to a symbiotic relationship that contributes to the effectiveness of digilantism in combating online fraud.

However, there are also serious ethical and legal concerns that are being posed, particularly in relation to the implications of assuming a law enforcement role without official sanction, and in relation to the nature of activities scambaiters engage in, which occasionally finds them acting against the law. Nonetheless, the collective effort of scambaiting communities provides a significant contribution to informal policing approaches, filling in gaps left by traditional law enforcement in combatting the rapidly evolving landscape of cybercrime (Wall, 2024).

The role of the internet as a technological enabler in digilantism is as central as it is multifaceted. Primarily the internet provides anonymity and greatly expands an individual's reach, allowing them to undertake actions against perceived injustice without the immediate risks that might be encountered in physical interventions (Trottier, 2017). Furthermore, the internet also offers tools and platforms that can amplify the actions of netizens, such as social media platforms, forums, as well as video hosting and streaming platforms. Such tools facilitate the aggregation and dissemination of data across wide networks in real-time, which, in the context of scambaiting, can be critical in identifying, disrupting, and exposing scamming operations. Furthermore, such platforms also enable collaboration between netizens, allowing them to share information, discuss strategies and tactics, and disseminate warnings to broader communities in an effective and efficient manner.

The collaborative nature of the internet affords collective intelligence and crowdsourcing which are particularly effective when it comes to scambaiting. Communities of netizens are able to pool their resources (time, expertise, and even financial) in order to identify, evaluate, and target scamming operations, but they are also able to utilise technological solutions to gather evidence and to submit this to appropriate law enforcement (or private entities in some cases) in order to enable them to take action and prevent scamming operations from coming to fruition. Collecting information about scammers' identities, locations, bank accounts, etc., and passing it to the appropriate entity who is able to act on such information is a frequent occurrence in scambaiting operations.

The evolution of the internet and of communication technologies over the past few decades have fundamentally altered the dynamics at play in scambaiting, by enabling scambaiting operations to increase their visibility and impact through leveraging the latest technologies and platforms. Previously isolated efforts of lone actors now benefit from channels that allow them to gain viral attention, rallying extensive support and leading to more significant disruptions to scamming operations. An apt example of this consists of the increase usage of YouTube and other social media platforms for documenting and disseminating scamming strategies in order to provide educational material to the wider public, but also to provide an additional deterrent

to potential scammers who might are now facing the risk of public exposure and ridicule (Cross & Mayers, 2021).

Furthermore, the capacity held by the Internet in terms of rapid dissemination of information enables successful tactics and strategies to be shared and replicated across multiple contexts and jurisdictions, contributing to the global resilience against online fraud. The collective action, enabled by digital connectivity, can be viewed as a form of social immunization against scams where informed netizens are able to act as both shields and educators for the less informed, thereby reducing the overall susceptibility of the online community to fraud. However, it is important to mention that the effectiveness of dissemination can be disrupted by language barriers and access to appropriate platforms, as well as by levels of technical proficiency. Additionally, while scambaiting contributes to the development of a wider resistance to online fraud, scamming operations are seeing similar evolutions to leverage technological advancements and new platforms to reach potential victims.

One of the most notable technological advances in recent years has been the emergence of AI technology in the form of deepfakes (hyper-realistic digital forgeries of images or videos that are generated via AI technology) and LLMs (large language models that can generate coherent and contextually appropriate text based on user-provided prompts). The adoption of such technologies has been quite rapid among both online scammers and online scambaiters, with social media platforms seeing their fair share of deepfakes impersonating various celebrities providing investing advice and directing users to illegitimate platforms designed to defraud them. The implications of deepfakes do however go beyond online fraud with Chesney and Citron (2019) discussing the detrimental effects of such technology at a macro-political scale.

1.7 The Interplay of Vigilantism and Law Enforcement

In recent years, vigilantism has emerged as a significant phenomenon that interacts with law enforcement in complex ways. The literature on this topic explores various aspects, including the factors that foster support for vigilantism, its trends and implications, and the challenges it poses for police forces.

One study conducted in Mexico suggests that citizens' support for vigilantism is influenced by their perceptions of community trust and the trustworthiness of law enforcement. When individuals perceive a strong sense of community trust coupled with distrust in law enforcement, they are more likely to support vigilantism (Zizumbo-Colunga, 2017). The degree to which people have confidence in authorities influences how social capital is utilised and displayed, impacting citizens’ attitudes towards vigilantism. In instances where there is higher

trust in authorities, engagement in or support of vigilantism is less likely and conversely, lower trust in authorities might lead to higher participation in vigilantism as people feel they need to take matters into their own hands.

In Ghana, vigilantism has been on the rise as a means of self-defence, enacting justice, and sanctioning wrongdoings, with most cases reported in urban areas. The primary suspected crimes leading to vigilantism are theft and robbery, with victims consisting typically of males between the ages of 20 and 45. Factors contributing to this rise in vigilantism include the perceived corruption, low responsiveness, mistrust, and ineffectiveness of the police and justice system. This trend raises concerns about the disregard for human life and dignity and highlights the need for a re-evaluation of justice administration systems and the empowerment of law enforcement to combat crime effectively (Adzimah-Alade et al. 2020).

Nigeria presents another instance of vigilantism being employed to address security issues due to inadequate law enforcement provisions. Vigilantism has been observed to emerge in communities in Anambra State with factors such as age, education, fear of crime, and distrust in the police influencing individuals to participate in vigilante groups aiming to promote fairness, justice, and accountability in society (Anadi, 2017). Despite traditional community-based policing having been replaced by a modern police force during colonialism times in Nigeria, this marginalized traditional indigenous communal policing leading to a rise in vigilantism as a self-help measure in the context of inadequate modern policing. The relationship between official state police forces and vigilantism in Nigeria appears to be complex, having significant impacts upon security and governance (Baasiru & Osunkoya, 2019).

Law enforcement agencies face a dilemma when dealing with vigilantism. While some police officers may empathize with the motivations behind vigilantism, they are also tasked with upholding the law, which vigilantes often break. The media attention that vigilante incidents attract can lead to intense scrutiny of police practices and their responses to such incidents (Silke, 2001).

Even more interestingly, law enforcement agents have also been observed to engage in vigilantism in instances where the constraints of the legal framework they are operating has been deemed insufficient towards maintaining law and order. Jauregui (2015) touches upon the practice of “encounter killings” carried out by police officers in India, consisting of planned operations where suspected criminals are targeted under self-defence claims. Cultural elements and media portrayal contribute to an apparent tacit public support of such activities but the glorification of such extralegal violence can distract from addressing systemic issues that enable it such as corruption and inadequate law enforcement provisions.

In Russia, vigilante justice has evolved with some groups cooperating with the police while others have conflictual relations. Vigilante groups have been observed leveraging social media platforms towards garnering support and towards publicizing and monetising their activities by distributing footage of engaging in vigilantism. Furthermore, an interesting relationship is revealed between vigilante groups and the Russian state, depending on their alignment. Vigilante groups are not homogenous, with some of opposing the Russian government, while others are acting in favour of it, even receiving funding from it. (Favarel-Garrigues, 2021).

The phenomenon of cybersecurity vigilantes has also been examined, with Silva discussing the detrimental effects of vigilantism in the realm of cybersecurity as it pertains to cybercrime investigations. Vigilantism has been noted to lead to loss or alteration of digital evidence, thus obstructing law enforcement efforts to tackle cybercrime through legitimate channels. However, recognition is also given to how the expertise that vigilante actors sometimes possess can be useful to combating cybercrime. Such actions reflect a shift in how law enforcement is conducted online and may shape the future of cooperative criminal justice. (Silva, 2018).

Vigilantism is a multifaceted issue that challenges the traditional roles of law enforcement. It arises from a combination of community dynamics, perceived inadequacies in the justice system, and the public's response to crime. The interaction between vigilantism and law enforcement is characterized by a delicate balance between understanding the motivations behind such acts and the necessity to uphold the rule of law.

1.8 Introduction to Scambaiting and Online Fraud

One particular form of digilantism that we will focus on is scambaiting, which involves individuals, often referred to as scambaiters, that engage online scammers with the intent of wasting their time and resources and preventing them from victimizing others. The practice of scambaiting has seen significant development alongside the evolution of internet and communication technologies and it is a unique point at the intersection between cybersecurity, ethics, and social activism.

Fraud is not a new phenomenon by any means and history is filled with examples of fraudulent activities being carried out. The UK Fraud Act 2006 makes the distinction between three types of fraud: (1) by false representation; (2) by failing to disclose information and; (3) by abuse of position. Martin (2003) defines fraud as “A false representation by means of a statement or conduct made knowingly or recklessly in order to gain material advantage”. As such, frauds are depriving individuals or companies of money/property/rights by means of deception and misinformation. Deception is the central element employed in committing fraud and convincing

a victim to willingly part with their money, rights, or possessions. Fried (2001) argues that the internet can be viewed as a breeding ground for fraud.

In recent years, online scams have become a ubiquitous element of modern life on a global level (Button, 2012; Levi, 2008; Smith, 2010). Scam emails and phone calls are prevalent in the contemporaneous world, and it would be challenging to find someone who has never been the target of such fraudulent activities. The expansion of internet communication technologies has significantly lowered barriers for fraudsters, enabling them to target a vast global pool of potential victims at scale (WEF, 2023; ALI, 2019). Simultaneously, rapid technological advancements have yielded an increasingly complex digital environment, defined by sophisticated e-commerce interfaces, algorithmic manipulation, and technical jargon, that many users find difficult to navigate safely (OECD, 2023). As such, online scammers who aim to take advantage of those who are less computer savvy might find an increasing number of people falling in this category. However, technology has not only contributed to the expansion of scamming activities across borders and IT systems. Various technological solutions have also been developed and continue to be developed to combat illegal online activities. Examples of these include antiviruses, email filters, encryption, 2-factor- authentication, and many others. These measures are, however, predominantly defensive and reactionary in their development such that online scammers have learned, and continue to learn, how to by-pass such tools or even turn them to their advantage. Advancements in technology are also picked up by scammers to further improve their schemes and strategies. For example, encryption is now being used in ransomware (online viruses that encrypt personal files holding them ransom in exchange for payment), and 2-factor-authentication can be by-passed via social engineering. Furthermore, increasingly complex technological solutions to the problem of online scamming also led to an increase in the complexity of online environments and those who are not adept at keeping up with most recent technological advancements may continue to remain susceptible to scamming attempts.

Across the UK, fraud accounts for over 40% of all crime, with over 80% of fraud being cyber-enabled, and over 70% either originating online or involving an online element (Crown Prosecution Service, 2023). Research commissioned by Ofcom has found that approximately 9 in 10 adults in the UK (87%) have come across online content they suspected to be a scam or fraud (Ofcom, 2023). Research by the National Trading Standards has found that 73% of adults in the UK have been targeted by scams, with 35% having lost money as a result. Additionally, fewer than a third of these have reported the crime to an authority, out of which 47% reported having been made to feel stupid or embarrassed, 34% reported feeling heard and understood, and just 38% felt strongly that their case was taken seriously (National Trading Standards, 2023).

In the US, the Federal Trade Commission (FTC) reported that fraud losses totalled almost $8.8 billion in 2022, having increased by 30% from the previous year, with imposter scams being the most prevalent, accounting for the highest number of reports (FTC, 2023). Separately, the FBI Internet Crime Report (2023) has found that more than $12.5 billion was lost in 2023 to online fraud, which is a 22% increase from the previous year. An Ipsos poll conducted in 2023 on behalf of Wells Fargo found that nearly 1 in 3 Americans (31%) have been the victim of online financial fraud or cybercrime. A Europol Spotlight Report on Online Fraud (2023) noted that online fraud schemes represent a major crime threat in the EU and beyond, with online fraudsters generating billions in profits every year (Europol, 2023).

One prevalent scam, the Nigerian 419 scam, is the digital adaptation of a historical advance-fee fraud dating back centuries, known as the “Spanish Prisoner”. As part of this scheme, the fraudster would engage their target under the pretence of being an exiled Spanish noble who has been forced to flee his country by various circumstances. The fraudster would then claim that he has a sister who remains imprisoned back in Spain and that he requires assistance in arranging her rescue, especially since he had no resources or money available. The target is then informed that the sister is in possession of various riches and, upon successfully becoming free, the fraudster and his sister would be willing to reward the target for the assistance provided. If convinced by this narrative, the target will provide the fraudster with the requested funds for arranging the rescue of the captive sister. Subsequently, the fraudster might return to inform the target of additional trials and tribulations that have emerged which now require further financial contributions, in an attempt to extract even more money from the target. Potentially, the target will continue to provide funds to the fraudster until such a point where either the target realises what has happened or they run out of funds and the fraudster breaks contact, moving on to other targets (Yar and Steinmetz, 2019). In present day, this particular scam is more commonly known by a different name, the ‘419 scam’, named so after the relevant Nigerian penal code prohibiting this activity.

In the Nigerian 419 scam, the victim is most commonly contacted by the fraudster via email (although older instances would have also seen physical letters being utilised) who claims to be the holder of a high governmental position, or a member of some royal family. The fraudster then proceeds to claim that he is in the possession of significant riches but is unable to access them due to various unfavourable conditions. He requests the assistance of his target who needs to provide the fraudster with a certain amount of money to be used in releasing the fraudster’s alleged riches. In exchange, the fraudster expresses his intention to reward the target with a portion of the riches (which usually consists of significant amounts of money). This scheme is also commonly known as ‘advance fee fraud’ due to the strategy of the fraudster attempting to extract a fee from the target in advance of providing them with untold riches.

There are numerous variations to this scam, including lottery winnings that need an advance fee in order to be claimed or a surprise inheritance from a very distant relative that will be sent to the target once they cover various fees.

Another common type of fraud in the modern day consists of romance scams, which have become increasingly commonplace. These fraudulent schemes exploit the increasing number of dating platforms available online and aim to get access to their victims via pretending to be romantically interested in them. Once contact has been established by the fraudster with their target and the fraudster has been relatively successful in building a relationship with the target, there are multiple ways in which the fraudster might attempt to extract funds from their victim. Commonly, it is likely that the scammer would contact their victim under the pretence of sudden significant hardships that have affected them (health issues, legal problems, unexpected expenses), and they appeal to the victim for financial aid in resolving these issues. Frequently, the victim is placed in a position where not assisting the fraudster financially might come across as their feelings not having been genuine, and fraudsters utilise this, often leveraging the romantic emotions of the victim towards extracting funds out of them so that the relationship may continue. Whitty and Buchanan (2012) found that individuals who are high in romantic beliefs have a higher tendency towards falling prey to such scams. If successful, this scheme can be continued by the fraudster until not more funds can be extracted or until the victim realises what is happening and breaks away from the hold of the scammer.

Romance scams are particularly problematic in the sense that it is not solely the financial loss that affects the victim; the emotional toll of losing the romantic relationship is also significant. The victims can suffer significant emotional distress and can be plagued by feelings of betrayal but also grieving the loss of the relationship (Button and Cross, 2017; Whitty and Buchanan, 2016). Despite the fraudulent nature of the scam, the feelings experienced and expressed by the victim are often genuine and as such, the termination of the relationship has a significant impact upon them. This distress is often amplified by the fact that the trauma is often not understood by the victim's close ones (Whitty and Buchanan, 2016). Individuals might often be reluctant to report having fallen victim to a crime due to feelings of shame and embarrassment. In cases of romance scams these feelings can be amplified due to the significant emotional dynamics that are at play, which Walker (2011) uses to argue the possibility that such offences are even more commonly unreported.

Button, et al. (2014) found that some of the reasons for which people fall victims to online scams include the diversity of fraudulent schemes, the relatively low amounts of money sought by perpetrators, grooming, and pressure and coercion employed by scammers. Looking at the victims of romance scams, Saad (2018) found that individuals between the ages of 25 and 45

years were more likely to be victims of romance scams in Malaysia. Furthermore, the majority of the victims had completed some form of education, were in employment, and had a monthly income. Only 17% of the respondents to the survey who were victims of romance scams were not earning an income. Additionally, both single and married individuals were found to be at risk of becoming victims of romance scams. Lack of computer skills and limited awareness of online scamming was found to increase the likelihood of becoming a victim. Whitty (2018) found that middle-aged people are more likely to become victims of romance scams than younger or older people and that, contrary to what was predicted, those who were more highly educated were also more vulnerable to this type of fraud. Fisher, Lea, and Evans (2013) suggested that overconfidence could lead to individuals becoming more vulnerable to scams and that a higher level of education might be more likely to cause overconfidence.

However, Lazarus et al. (2023) points out that, while multiple pieces of research seek to tackle the issue of romance scams, the majority of the studies focus on victim perspectives with insights on offender motivations, methods, and socio-cultural contexts remaining limited. Similar issues can be observed more broadly within the research on online scams, where methodologies employed may fail to capture offenders' perspectives, which create a gap in the academic understanding of online fraud.

From the perspective of law enforcement, online scamming is a challenging issue to tackle for a number of reasons. First, the relatively borderless nature of the internet makes it very easy for scammers to access victims from all around the globe, not being restricted to particular geographical areas of activity. This is problematic from a law enforcement point of view because law enforcement agencies generally operate in a compartmentalized and hierarchical manner, based primarily on geographical areas. In events where an online scam is reported, it is often the case that the perpetrator would have operated outside the reach of law enforcement that is local to the victim (generally the perpetrator could even be based in a different country).

Second, the nature of scamming often leaves victims with a sense of shame and embarrassment that they have been tricked in such a manner. These feelings often lead the victim to not report the crime for fear of their perceived gullibility to be further exposed, instead attempting to move on past this event and keep it hidden so as to save face. This train of thought is prevalent not only in cases of individuals falling victims to fraud but also in cases of organizations being affected by fraudulent activities. It is often the case that, upon fraud being uncovered within an organization, the preference is for this to be dealt internally and kept from the eyes of the world for fear of the reputational damage that the organization might further suffer. Since the crime is not reported to authorities, the perpetrator can simply move on to another target. Additionally, victims of scams might also be concerned about having to deal with

the complexities of the legal system if they chose to report the crime, concern which could deter them from reporting the incident.

Third, with online scamming being so widespread and common, it would not be viable from the perspective of law enforcement resources to extensively investigate and tackle each incident. Furthermore, considering the fact that many of these investigations would have a low chance of successful apprehension of the offender due to factors explored above, and the toll that the investigation could take on the victim (in addition to what was suffered due to the scam), it might even be considered counterproductive to expand resources on such investigations.

The confluence of motivated scambaiters, nimble scam operations, variable institutional responses, and engaged civil society has given rise to the concept of 'scambaiting', the act of feigning victimhood to waste scammers' time. From a theoretical standpoint, this phenomenon can be analysed through Young's Square of Crime, which highlights how interactions among offender, victim, state, and society shape crime dynamics (Young, 1991). Scambaiting itself is increasingly recognized as a form of cyber vigilantism that mobilises civic actors to counter fraudsters (Bérney, Ondrus & Holzer, 2024), and is conceptualised within academic research as a social-practice of deception, entertainment, and informal crime disruption (Dynel, 2021). Through this lens, scambaiters insert themselves into the equation attempting to take the place of prospective victims. Considering that the primary goal of online scamming is financial or material gain, the intervention done by scambaiters reduces the amount of money that the perpetrators of online scams can gain while increasing the amount of time they spend with potential victims.

It is important to note that scambaiting works as an umbrella term, referring to a variety of activities that individuals engage in with the purpose of interfering with online scams. Furthermore, scambaiters are generally not affiliated with law enforcement agencies, and in cases where they might be affiliated, they do not conduct their scambaiting activities as part of their role within law enforcement. It is however worth noting that cooperation with law enforcement or private organizations is frequently sought by scambaiters in attempts to increase the effectiveness of their activities and to thwart the activities of scammers. It can be argued that scambaiting can be categorised as digilantism (a term sometimes used interchangeably with "netilantism"), which is a variation of vigilantism that occurs primarily in the digital arena.

Les Johnston (1996) has posited that there are six criteria that an activity needs to fulfil in order to be considered an act of vigilantism: (1) A certain level of planning, premeditation, or organization needs to occur from the individual(s) responsible for the act; (2) The act must be taken by private agents, unaffiliated with law enforcement (although an argument can be made

about those affiliated with law enforcement acting outside of this affiliation); (3) The act must be autonomous and free from state support, it should not be sanctioned by the government or by agencies in the criminal justice system; (4) The act must encompass the use of, or threatened use of force (Smallridge et al. (2016) argue that this criteria can be broadened to also include causation of harm in order to fit in with the more modern digilantism); (5) The vigilante activity needs to occur in response to a criminal or socially deviant event that transgresses institutionalized norms. This transgression can be either real or perceived; (6) The goal of the vigilante act must be to control crime or social deviant actions, specifically done by offering assurances of personal and collective security to people. Because scambaiting activities generally meet all of Johnston's criteria and there is a strong digital element to them, digilantism appears to be an appropriate category to be used.

1.9 Thesis Structure Overview

The chapter that follows aims to provide a comprehensive literature review that explores work already done on researching vigilantism and scambaiting, while also exploring significantly related topics such as broader cybercrime and online content monetisation strategies. The section also briefly touches upon criminological theories and theoretical frameworks that could offer a window for scambaiting to be examined through.

Next, the methodology chapter seeks to outline the details relating to the research design, touching upon data collection, analysis, theoretical frameworks, ethical considerations, as well as challenges and limitations. Following this, the discussion chapter aims to return to the topic of scambaiting in light of the literature reviewed and of the findings of the data collection and analysis. The first section seeks to revisit the topic of scambaiting as vigilantism, exploring motivations for scambaiting alongside comparison with traditional counter-fraud strategies and some of the notable ethical ramifications. Detailed attention is also given to the strategies and tactics employed by both scammers and scambaiters towards enabling an examination of the interplay between these two parties, as well as of some moral considerations that emerge. Considering the significant use of fictitious identities being adopted by both scammers and scambaiters, a section has been dedicated to exploring the identities being created, as well as how they are constructed and the implications this has for the understanding of victimhood. The emotional dynamics and psychological elements that can be observed in scambaiting interactions are touched upon next, followed by a detailed look at the role played by audiences within the perpetration of modern scambaiting strategies. The last two sections of the discussion chapter explore the use of monetisation strategies by scambaiters, and the use of

technological tools by both scammers and scambaiters, with the chapter's conclusion looking at future trends in scambaiting and future research.

1.10 Chapter Summary

This chapter has established the context for the thesis by outlining the significant and escalating threat of online fraud, a global issue causing substantial financial and psychological harm. In response to the perceived limitations of formal law enforcement in addressing this transnational and technologically sophisticated crime, a form of digital vigilantism known as scambaiting has emerged and grown in prominence. The chapter has argued that while this practice is increasingly visible, particularly due to content monetisation models on platforms like YouTube and Twitch, it remains significantly underexplored within academic literature.

To properly situate scambaiting, a historical analysis of policing in the UK was undertaken, tracing its evolution from informal, community-based systems like the 'watch and ward' to the establishment of a centralised, professionalised state police force. This historical trajectory, culminating in contemporary neoliberal approaches that involve privatisation and public-private partnerships, provides a crucial backdrop for understanding scambaiting as a modern form of informal justice that occupies the gaps left by traditional policing.

The theoretical concept of vigilantism was introduced, using Johnston's (1996) criteria to build a foundational understanding. This was then differentiated from its modern counterpart, digilantism, which leverages digital platforms to achieve similar goals but with distinct methods, scales, and impacts. The chapter positions scambaiting as a specific and highly relevant manifestation of digilantism, driven by 'netizens', both active participants who directly engage scammers and passive audiences who provide support and legitimacy.

Finally, the chapter outlined the complex and often fraught relationship between vigilantism and formal law enforcement, highlighting how citizen-led justice initiatives arise from a lack of trust or perceived ineffectiveness of official channels. By framing scambaiting within these historical, theoretical, and socio-legal contexts, this introduction has laid the groundwork for the thesis's core objectives: to analyse the practices, motivations, and implications of modern scambaiting, with a particular focus on the impact of content monetisation, through the theoretical lenses of dramaturgical analysis and symbolic interactionism. The subsequent chapters will build upon this foundation to provide a comprehensive academic examination of this complex phenomenon.

Chapter 2 Literature Review

2.1 Introduction

The rapid advancement of digital technologies has not only transformed traditional social interactions but has also given rise to novel forms of crime and corresponding countermeasures, among which we can count scambaiting. However, there is a significant academic debate regarding the novelty of such crime. Some scholars argue that much of what is labelled 'cybercrime' is effectively 'old wine in new bottles', representing traditional offences merely displaced into a digital environment (Grabosky, 2016). In contrast, others contend that the digital context creates genuinely new criminal opportunities and methods (Wall, 2007). A useful distinction to navigate this debate is between 'cyber-dependent' crimes, which can only exist in a digital context (e.g., hacking, malware distribution), and 'cyber-enabled' crimes, which are traditional offences amplified in scale or reach by the internet, such as online fraud (McGuire and Dowling, 2013). Scambaiting, as a response to the latter, can be categorised as a form of digilantism, and it involves individuals (or communities of individuals) engaging with scammers with the goal of wasting their time, disrupting their operations, gathering information and potentially protecting vulnerable targets from falling victims to online fraud. The practice of scambaiting is situated at the intersection of technology, law, ethics, and social activism, making it an appealing topic for academic research under the broader umbrella of criminology.

The aim of this literature review is to contextualise the phenomenon of scambaiting within the larger discourse on cybercrime and digilantism, seeking to explore both the historical roots, as well as contemporary manifestations of scambaiting activities. Through examining academic literature, media reports, case studies, and other materials, this review aims to delineate the evolution of vigilantism into its digital form and to observe the implications of such transitions on legal frameworks, societal norms, and ethical considerations.

Key topics covered include traditional forms of vigilantism and the transition into digilantism, previous academic studies on scambaiting and related cybercrimes, and the economic models of online streaming and content monetisation, especially as pertaining to the most recent scambaiting strategies and models. Through synthesising diverse perspectives from criminology, sociology, information technology, and media studies, this literature review hopes to provide a comprehensive overview of the field surrounding scambaiting, and to highlight areas where further research is needed.

Considering the multifaceted natures of both scambaiting and the online fraud 'ecosystem', that is, the complex and interdependent network of actors (fraudsters, victims, scambaiters, law enforcement), technologies (platforms, payment systems), and economic incentives that facilitate and sustain online fraud, this review also aims to address the ethical and legal challenges posed by some of the activities that scambaiters engage in. Critical questions about the legitimacy and consequences of scambaiting are central to the topic and providing adequate consideration to both potential benefits and risks of combating online fraud through unauthorized forms of justice is necessary towards appropriately tackling the subject of scambaiting.

In summarizing the existing body of knowledge, this literature review sets the stage for subsequent empirical research, guiding the thesis towards addressing unexplored or underexplored aspects of scambaiting, with the goal of contributing to both theoretical and practical understanding of how this form of digilantism fits within the wider ecosystem of cybersecurity and law enforcement strategies, ultimately contributing to a safer digital landscape.

2.2 Theoretical Frameworks for Understanding Scambaiting

To understand the complex phenomenon of scambaiting, it is necessary to draw upon a range of theoretical frameworks from sociology, criminology, and systems science. These theories provide a foundation for understanding the motivations behind vigilantism, the structure of the online fraud environment, and the specific dynamics of the interactions that define scambaiting.

In exploring the theoretical foundations of vigilantism, it is useful to draw on sociological theories, particularly those of Max Weber and Emile Durkheim, in order to understand how elements of social structures and collective sentiment can feed into the emergence of vigilante practices and activities.

Max Weber's theory of authority plays an important role in understanding the rise of vigilantism. Weber distinguishes between three types of authority: traditional, charismatic, and legal-rational (Weber, 2014). The emergence of vigilantism can often occur in circumstances where the legal-rational authority… is weak or perceived as illegitimate. This theoretical perspective is supported by empirical research, which has consistently shown a strong correlation between low public confidence in police responsiveness and higher levels of support for vigilantism (Haas et al., 2014). This is particularly relevant in the context of online fraud, where the police response has been identified as often inadequate, creating a vacuum that citizen-led initiatives

like scambaiting may seek to fill (Cross and Blackshaw, 2015). Under such conditions, individuals or groups may revert to what Weber might categorise as a form of charismatic authority, where justification for action is based on the perceived extraordinary qualities of the group or its leaders, or even a return to traditional authority, where-established customs guide justice rather than formal laws.

Weber also made the argument that a state is assuming monopoly on the legitimate use of physical force (Anter, 2019). In instances where this monopoly is contested or when it is found to be inadequate or insufficient, alternative forms of authority may emerge. In instances where members of a community might perceive the legal system as corrupt or inefficient, they might feel justified in bypassing formal legal structures and carrying out justice via informal means, thus giving rise to vigilante activity. Empirical studies, such as Badiora's (2019) work on community support for vigilantism in Nigeria, demonstrate how these perceptions of state failure directly shape and legitimize community-led justice efforts.

Durkheim's concepts of collective consciousness and social cohesion are also pivotal. According to Durkheim, a society's collective consciousness includes the shared beliefs and moral attitudes which operate as a unifying force within society (Smith, 2014). Durkheim saw law as an expression of collective consciousness, but doubts exist about the possibility of this occurring in modern society which is vulnerable to democratic crises (Greenhouse, 2018). In instances where formal mechanisms fail to enforce law and order effectively, a disconnect occurs between collective consciousness and formal law enforcement channels, which may prompt members of the community to engage in vigilantism in order to restore moral order and maintain social cohesion. Such emergence of vigilantism can be described as "normative" due to involving action from community members with the goal of defending their shared norms and values.

The concept of anomie can also contribute to the understanding of the circumstances that give rise to vigilantism. Durkheim introduced the term "anomie" to describe a state of normlessness, where there is a lack of commitment to shared values, standards, and rules necessary to regulate individual behaviours and aspirations within the encompassing social environment. This can occur as a result of rapid social change and weakening of traditional institutions (Bernburg, 2019). During times of rapid social change or stress, the collective consciousness might weaken, leading to the emergence of a state of anomie where norms are confused, unclear, or absent. This state of normlessness can give rise to higher rates of crime and deviance, which could encompass forms of vigilantism which might arise as an attempt to re-establish clear norms and order when the state is unable to enforce laws effectively or when laws are no longer appropriately reflecting the community's moral convictions.

Robert K. Merton's further development of the concept of anomie provides a further link between social structures and vigilantism (as deviation). Merton describes anomie as emerging out of a disconnect between socially approved goals and the availability of legitimate means to achieve these goals (Featherstone, 2003). Where societal goals such as justice and security are out of reach for the members of a community through legitimate channels, alternative strategies could be sought which can include vigilantism as an innovative response within this framework in instances where the goals sought consist of the establishment and maintenance of law and order.

While both Merton and Durkheim discuss anomie in the context of social structure and its effects on individual behaviour, there are significant differences in their approaches to the concept. Durkheim's theory is more concerned with individual self-interest becoming increasingly unrestrained due to the flourishing of a liberal free market ethos, arguing that such an economic system can lead to a decrease in social cohesion and collective conscience. When individual values are prioritised over communal values, anomie can emerge leading to social instability and personal disorientation. Merton's approach to anomie focuses more on the disjunction between cultural goals and the institutionalized means available to achieve them, causing strain which can manifest in individuals stepping outside of legitimate and approved channels in order to achieve the desired goals. Despite vigilantism being more concerned with concepts of fairness, justice, and social order than achievement of goals, a relationship can be drawn to the concept of anomie. Interestingly, in Durkheim's view of decreasing social cohesion and collective conscience in the context of increasing emphasis on liberal free market approaches to social structures, vigilantism (especially as enabled by free market channels) can contribute to the alignment of individual self-interest (of the perpetrators of vigilantism) with the interest of a wider social group (united in opposition to the target of vigilante activity). Furthermore, in Merton's strain theory viewing anomie as arising from the insufficiency of legitimate channels to achieving cultural goals, vigilantism can be a result of such anomie in instances where the goal sought is establishment or maintenance of social order in the absence of appropriate sanctioned law enforcement solutions.

Merton's strain theory has been taken further by Agnew who developed General strain theory, expanding into considering a broader range of strains and factors that can influence the likelihood of coping through illegitimate or illegal solutions. In addition to Merton's emphasis on inability of achieving cultural goals due to insufficient means, Agnew also considers strains relating to personal relationships and individual circumstances, placing significant emphasis on the emotional responses to strain. The manner in which individuals cope with and respond to strains is theorized to affect how they resort to criminal behaviour (Agnew, 2002). This expansion of strain theory better situates vigilantism because its emphasis on the emotional

responses to strain, such as anger, frustration, and a sense of injustice, directly corresponds to the motivations often expressed by scambaiters, who frame their actions as a necessary response to the harm and emotional distress caused by scammers. It accounts not only for the actions of those who actively engage in vigilante activity, but also extends to those who support and encourage it.

Another theoretical framework that can be used to situate vigilantism is Cohen and Felson's Routine Activity Theory (RAT), which argues that, for a crime to occur, three elements need to converge in space and time: a motivated offender, a suitable target, and the absence of a capable guardian (Cohen & Felson, 2010). While RAT was initially developed with consideration for offences occurring in a physical environment, its framework can also extend to cybercrime (Leukfeldt, 2016). The principles of RAT extend to the temporal and spatial dimensions of online activities, suggesting that the timing and location of cyber intrusions align with the convergence of a motivated offender, suitable target, and absence of capable guardian (Bock et al., 2017). Furthermore, empirical assessments of RAT in the context of cybercrime have found support for its components, demonstrating patterns of victimization that align with the theory's predictions (Choi, 2008). Kigerl (2012) also discusses how wealthier nations having more internet users per capita experience higher cybercrime activity, which can be further amplified by higher unemployment rates. Additionally, significant changes in opportunity situations, such as the increased number of online users during the Covid19 pandemic, has been found to correspond to a surge in cybercrime, supporting the RAT perspective that changes in routine activities can escalate the convergence of the motivated offenders and suitable targets in the absence of appropriate guardianship (Govender et al., 2021).

In the context of vigilantism, a key driver for vigilante activity is often the perceived absence of appropriate guardianship, not in relation to the vigilante activities themselves, but rather in relation to the targets of vigilantism, which could encourage individuals to take law enforcement in their own hands, thus stepping in to at least partially fill in for the absenteeism of capable guardianship. From this perspective, vigilantism can be seen as a means of compensating the absence of appropriate guardianship, attempting to compensate for the shortcomings (or absence) of appropriate law enforcement solutions, disrupting the convergence of motivated offender and suitable target, or retaliating against the motivated offender to prevent further reconvergence with suitable targets.

Social Control Theory is another significant paradigm in the field of criminology that looks at how society influences individuals to conform to laws and societal norms. The theory posits that people's relationships, commitments, values, norms, and beliefs encourage them to not break the law. Therefore, if moral codes are internalized and individuals are tied into and have a stake

in their wider community, they will voluntarily limit their propensity to commit deviant acts (Orcutt, 2016). The origins of social control theory can be traced back to the works of Travis Hirschi, who formulated it in the 1960s, emphasizing the role of social bonds in preventing crime. Four elements were identified by Hirschi as being related to the idea of social bond: attachment, commitment, involvement, and belief, being critical in fostering conformity and discouraging deviant or criminal behaviour (Muhammad, 2015). Robert Sampson and John Laub have later expanded social control theories to also posit that an individual's bonds to society can change over time, thus influencing their propensity towards deviant behaviour at different stages of life. Similarly, societal changes can also impact an individual's bond to society, having the potential to either weaken or strengthen these bonds, subsequently influencing an individual's predisposition to engage in deviant activities (Costello, 2017). Societal processes of change (sociocultural, political, and economic) have been gaining prominence as agents of change, shaping the societal framework and potentially altering mechanisms of social control (Rabie, 2013).

In relation to vigilantism, the weakening, breaking down, or complete absence of social bonds can leave individuals feeling disconnected or disillusioned with societal mechanisms and structures. In contexts where individuals perceive that the formal mechanisms of social control are inadequate or unjust, they might be incentivized to take justice in their own hands. Vigilantism can as such be seen as a response to perceived failure of social controls if the formal systems are not providing safety or justice. Individuals might feel compelled to step outside the prescribed systems and enforce norms and laws as they see them, especially if they do not believe that the law or its enforcers are legitimate or effective. On the other hand, vigilantism can give rise to the formation of new social bonds, particularly within vigilante groups and networks that have shared beliefs about justice and order. This can subsequently reinforce individuals' commitment to vigilante actions with the group itself substituting the conventional societal bonds that are perceived as insufficient.

Vigilantism can also emerge in contexts where there is ambiguity or conflict regarding social norms and values. In instances where societies are undergoing rapid change or experiencing cultural conflicts, norms might not be universally accepted or clearly defined, resulting in a fragmentation of the society in question. The ambiguity coming with this fragmentation can weaken the belief component of social bonds, leading to differing interpretations of what actions and behaviours are justified. In such instances, groups engaging in vigilantism might view themselves as upholding the "true" norms and values of society, in an attempt to return to and preserve previously accepted norms, especially in instances where they perceive the state to be failing in its duties to enforce such norms.

It is however important to note that the social control theory does not appear to fully account for the proactive nature of vigilantism, with individuals not only breaking away from societal norms due to weakened bonds but also seeking to enforce their own interpretation of these norms.

We can see through the lenses of various theoretical approaches that vigilantism as a phenomenon can be described and understood via different frameworks, betraying its versatility as a concept but also its dependency on the specific socio-politico-cultural circumstances it emerges within. Vigilantism itself has the potential to encompass a broad range of activities, some of which might be viewed more positively or negatively, but all of which are closely connected to the social fabric of the corresponding communities. As such, it is no surprise that, as channels for social organization evolve via advances in communication technologies, a similar transformation can also be observed within vigilantism itself, adapting to leverage the most recent tools and platforms that are available within social groups that are becoming increasingly shaped and influenced by the digital space.

Beyond classical criminological theories, two additional frameworks are essential for analysing the specific dynamics of scambaiting: Symbolic Interactionism, with a specific focus on Goffman's dramaturgical analysis.

Symbolic Interactionism is a sociological perspective that explores how individuals create and maintain society through meaningful interactions. The emphasis is placed on the subjective meanings that people impose on objects, events, and behaviours, and how these meanings are constructed through social interaction.

Rooted in the work of George Herbert Mead and further developed by scholars from the Chicago School, Iowa School, and Indiana School, symbolic interactionism offers distinct methodological approaches to studying social interactions (Carter & Fuller, 2016). The Chicago School is often considered the origin of symbolic interactionism, rooted in the work of George Herbert Mead and further developed by his student, Herbert Blumer. This approach is characterized by strong emphasis on empirical research, particularly through qualitative methods such as ethnography and participant observation. Emphasis is placed on studying people in their natural environments to understand how they create and interpret symbols in everyday interactions, with many studies focusing on urban settings, examining diverse social issues (Blumer, 2012).

The Iowa School, associated with scholars such as Manford H. Kuhn, approaches symbolic interactionism slightly differently, by emphasizing quantifiable and somewhat more structured methodologies compared to the Chicago School. Quantitative measures such as psychometric tests and structured surveys are introduced to focus on how individuals assume roles based on

social expectations and the meanings these roles carry, often in the context of experimental and survey research.

Shifting the focus on structural aspects of society, rather than actions and activities of individuals, the Indiana School, closely associated with the work of Sheldon Stryker, approaches symbolic interactionism to look at how individual identities are shaped by social structures. Methodological emphasis is placed on identity theory, with the focus on how the structure of society impacts the identity of the individual, exploring the interplay between individual agency and social structure. Longitudinal studies might be preferred, aiming to look at changes in identity and interaction over time, providing insights into how social structures and individual actions influence each other dynamically. A mixed-method approach is often employed by the Indiana School, integrating qualitative depth with quantitative rigour to understand the complexities of social interactions.

Within this broader symbolic interactionist tradition, Goffman's dramaturgical theoretical framework offers a particularly valuable lens for understanding scambaiting. Dramaturgical analysis conceptualises social interaction through the metaphor of theatrical performance, emphasising the ways in which individuals present themselves to others in everyday life (Goffman, 1959). It focuses attention on the broader interactional setting, advocating for a holistic understanding of encounters as coordinated performances in which meaning is collaboratively constructed (Branaman, 1997). Social life is understood as organised into performances occurring on metaphorical "stages," where roles, scripts, and audiences interact to shape outcomes. These performances produce emergent definitions of the situation that cannot be reduced to the intentions or actions of a single participant (Goffman, 1959; Edgley, 2013).

Performances are sustained through boundaries, both spatial and symbolic, that separate the "front stage" of public presentation from the "back stage" of preparation and candid behaviour, boundaries maintained in order to preserve desired impressions (Manning, 1992). Dramaturgical analysis also identifies mechanisms such as impression management, role performance, audience segregation, and "face-work" that prevent disruptions and sustain interactional order (Goffman, 1967). These concepts are particularly relevant in scambaiting, where both scammers and scambaiters carefully construct and maintain fictitious identities in a competitive, deceptive environment.

The strategies and channels of distribution employed by modern scambaiting allow access to a significant number of interactions captured by scambaiters in their attempts to disrupt scamming operations. These interactions can potentially provide valuable insights into how identities are negotiated and constructed by scammers in their attempts to defraud potential

victims, and by scambaiters in their attempts to disrupt scammers' activities. Furthermore, the frequent deceptive nature of these interactions enables deeper analytical possibilities. Scammers navigate these exchanges with the goal of convincing potential victims of their legitimacy and maintaining this façade for as long as possible to maximise gains. Similarly, scambaiters employ their own deceptive tactics to convince scammers of their legitimacy as victims, sustaining these fictitious roles to maximise disruption.

Despite interactions being played out primarily between scammer and scambaiter, the complexity of these exchanges and the layered deception involved also afford insights into other interactions such as scammer–victim, scammer–third party, and scammer–competing scammer. Additionally, depending on the scenario, further interactions may occur where scammers or scambaiters engage with banks, retail store representatives, or even real victims. Another important dimension is the interaction between scambaiters and their audiences, offering insights into the community element of scambaiting.

The interactions captured and distributed by scambaiters are often complex and multifaceted, made more challenging to analyse by the fact that both scambaiters and scammers are operating deceptively in an antagonistic fashion. The combined application of symbolic interactionism, dramaturgical analysis, and systems thinking has the potential to provide significant insights into the negotiation of identities, the management of impressions, and the emergent interactional order between the parties involved in scambaiting, as well as the broader ecosystem of online fraud.

Goffman's dramaturgical framework, as a specialised form of Symbolic Interactionism, provides a powerful lens for understanding the scambaiting ecosystem within the wider online fraud environment. By conceptualising social interaction as a series of staged performances, dramaturgical analysis reveals how scammers and scambaiters actively manage impressions, control information, and navigate shifting audience expectations (Goffman, 1959). The metaphor of performance captures the complexity of these interactions: scammers adopt persuasive "front stage" personas that present them as legitimate businesses or trustworthy individuals, while scambaiters craft convincing victim identities designed to elicit scammer trust and prolong engagement for disruptive purposes.

This perspective highlights the importance of role performance, audience segregation, and "back stage" preparation in sustaining these deceptive encounters. For scammers, back stage regions may involve coordination with co-conspirators, preparation of fraudulent scripts, and adaptation of tactics in response to resistance. For scambaiters, back stage work includes planning counter-narratives, rehearsing false identities, and coordinating with fellow activists or audiences. Both parties engage in impression management to maintain the coherence of their

respective performances, deploying verbal and non-verbal cues, narratives, and symbolic markers to reinforce credibility in the eyes of their audiences, whether those audiences are scam victims, scammer teams, or online communities following scambaiting campaigns (Goffman, 1967; Manning, 1992).

The application of Goffman's key dramaturgical concepts to scambaiting is particularly revealing:

- Front stage: This is the public-facing performance. For scammers, the front stage consists of scripted phone calls, persuasive emails, or fake websites designed to project professionalism and legitimacy. For scambaiters, it is the ongoing interaction with scammers, where every response is crafted to sustain the fictional victim identity.
- Back stage: Here, performers drop their public persona and prepare for the next act. Scammers may exchange tips in online forums, rehearse scripts, or share lists of potential targets. Scambaiters similarly retreat to private communication channels to plan storylines, share scammer intelligence, and review performance strategies with collaborators.
- Performance teams: Both scammers and scambaiters often operate in groups. Scammer teams may coordinate roles such as "closers," "tech support," or "money mules" to keep the illusion alive. Scambaiters form collaborative teams to develop intricate scams-within-scams, maintain multiple characters, and collectively prolong the engagement.
- Face-work: This refers to the tactics used to preserve one's social image and avoid embarrassment. Scammers engage in face-work when confronted with scepticism by quickly repairing the performance with reassuring details or alternative explanations. Scambaiters also use face-work to recover from slip-ups, improvising explanations to prevent scammers from realising they are the target.
- Audience segregation: Scammers strive to keep different audiences, victims, co-conspirators, law enforcement, from overlapping to protect their operations. Scambaiters likewise separate their scammer interactions from the public entertainment side of their work to prevent leaks that might compromise ongoing engagements.

Symbolic Interactionism more broadly situates these performances within the meaning-making processes that define the scambaiting environment. It directs attention to how identities, roles, and definitions of the situation emerge through interaction. Scammers and scambaiters alike draw on shared symbols, metaphors, and scripts to shape how their actions are interpreted by others. For example, scambaiters often frame their activities as activism, justice, or public

service, while scammers may justify their behaviour through narratives of necessity, entitlement, or entrepreneurialism. These interpretations are not static; they are constantly negotiated and reshaped through ongoing interaction, feedback from audiences, and shifts in the perceived legitimacy of the performance.

Together, dramaturgical analysis and Symbolic Interactionism offer a robust framework for analysing scambaiting not merely as a reaction to crime, but as a complex social phenomenon involving identity negotiation, role performance, and the collaborative construction of alternative forms of social order. This integrated approach enables a richer understanding of the motivations, strategies, and social dynamics shaping the scambaiting ecosystem, while also illuminating the broader cultural and interactional processes underpinning online fraud.

2.3 Scambaiting and Related Cybercrimes

Before delving into the specifics of scambaiting practices, it is crucial to understand the societal context of fraud victimization that motivates such actions. A foundational concept in victimology is Nils Christie's (1986) notion of the 'ideal victim'. The ideal victim is typically perceived as weak, blameless, and engaged in respectable activity when victimized by a "big and bad" offender who is unknown to them. This social construction means that individuals who do not fit this narrow stereotype often struggle to have their victim status recognized and may face blame for their own victimization.

Victims of online fraud frequently fall into this category of the 'non-ideal victim' (Cross, 2018a). They are often perceived as greedy, gullible, or somehow complicit in their own financial loss, which leads to a significant degree of victim blaming in public and institutional discourse (Cross, 2015). This societal tendency to deny or question their victim status can lead to secondary victimization, discouraging reporting and leaving victims isolated. The perceived failure of the criminal justice system to adequately support these 'non-ideal' victims and the prevalence of victim-blaming narratives create a fertile ground for alternative forms of justice, providing a key motivation for scambaiters who see themselves as champions for a misunderstood and underserved victim population.

The emergence of scambaiting can be attributed to the increase of online scams which have proliferated alongside the expansion of the internet and related technologies. Academic literature has framed this practice in various ways. For instance, Byrne (2013) analyses '419 Digilantes' through the lens of 'radical justice online,' highlighting how these actors operate on the frontiers of legality to enforce their own moral code. Similarly, Chang, Zhong, and Grabosky

(2016) conceptualise these activities as a form of 'citizen co-production of cyber security,' where self-help and vigilantism contribute to the broader goal of a safer digital environment. Scambaiters employ a variety of tactics and strategies to engage scammers, ranging from simple interactions aiming to waste their time, all the way to elaborate operations that can disrupt scamming operations and identify perpetrators who are referred to the appropriate law enforcement agencies. These interactions and operations are often shared on social media platforms and dedicated forums among vibrant communities that span across the globe (Sorell, 2019).

Understanding scambaiting is a worthwhile endeavour for several reasons relating to academic and practical domains. From the perspective of cybersecurity, scambaiting shines a light on the contribution of civilians to cyber defence and to building resilience against scamming operations, suggesting an area for potential formalisation and enhancement of public cybersecurity efforts (Choo, 2014). The range of activities scambaiters engage in also bring about significant legal and ethical considerations around the permissibility of deception and harassment, even against fraudsters. The interactions between scambaiters, scammers, and other parties are also insightful in relation to the dynamics of online interaction and deception, providing a window into scammer tactics and the public's engagement with such threats (Whitty, 2015). Finally, scambaiting can also be seen as a form of informal social control, relevant to theories of deterrence and to the sociology of deviance (Pratt, 2010).

The academic exploration of scambaiting is still developing with existing literature being scarce and often limited to descriptive and case study methodologies. More systematic analyses would be beneficial towards better assessing the effectiveness of scambaiting as a deterrent, understanding its societal impacts, and exploring its educational potential in raising awareness around cyber threats (Cross, 2021). Furthermore, such activities also have the potential to inform policy and to refine law enforcement strategies in the digital realm.

Scambaiting as a term has an umbrella function encompassing a wide range of actions and activities that scambaiters engage in with the purpose of making a negative impact upon various scamming operations. Primarily, in its simplest form, scambaiting involves directly engaging scammers via the appropriate communication channels (phone, email, instant messaging, social media, etc.) with the goal of wasting as much of their time and resources as possible so that these resources cannot be employed towards defrauding potential victims. Additionally, one element of scambaiting communities consists of sharing and discussing scambaiting attempts, serving to educate members on the inner workings of various scamming operations, and to warn the public about the dangers of online scams (Sorell, 2019).

The phenomenon of scambaiting can be traced back to the early days of the internet when email-based scams (most notably the “Nigerian Prince” scam, or 419 scams) began to proliferate. As these scams grew in sophistication and number, internet users started organizing informal networks to combat them with the first recorded instances of scambaiting emerging in online forums and websites through users sharing stories of engaging with scammers. Over time, these efforts became more organised with the creation of dedicated websites and communities that not only shared effective scambaiting strategies but also provided resources for potential scam victims (Cross, 2021).

It is important to note that, similarly to how online scamming operations vary in the strategies and tools employed, scambaiting activities are similarly varied in how scambaiters attempt to tackle the issue of online fraud. Zingerle (2014) conducted an analysis of scambaiting on online forums and identified seven categories of scambaiters, based on the activities they engaged in. Within the environment of online forums, each of these categories were determined by particular activities that the users would engage in, but also by the particular goal of these activities. As such, one of the categories found by Zingerle consisted of Scam Alerters who would engage in discovering scamming attempts and report them within online scambaiting communities, while another category consisted of Romance Scam Seekers who would focus only on romance scams and make themselves “available” as targets on various platforms, trying to waste scammers’ time. Other categories of scambaiters include Trophy Hunters (who aim to obtain “trophies” which is something either physical or virtual that the scambaiter obtains from the scammer); Website Reporters (who identify and report illegitimate websites within scambaiting communities); Bank Guards (who aim to obtain banking information relating to scamming operations and report them to the appropriate third-party); Safari Agents (aiming to get the scammer to physically travel to various locations as a way of wasting additional time); and Inbox Divers (social engineers who aim to gain access to scammers’ communication channels and warn potential victims or report ongoing criminal activities).

Zingerle's categorisation, based on observations from only two online communities, is likely not comprehensive. However, it serves as an excellent example of the varied nature of scambaiting activities. Furthermore, the category of activities that constitute scambaiting continues to be fluid as it evolves and expands alongside both scamming operations, as well as alongside technological advancements.

In the realm of cybersecurity scambaiting can be found to share some similarities with other activities that have been evolving and gaining in popularity in recent years. One of the more significant concerns in cybersecurity that have emerged with the advent of the internet and computer technology is “hacking” (Berney, Ondrus & Holzer, 2024). Similarly to scambaiting,

hacking acts as an umbrella term, encompassing a wide range of activities that aim to gain unauthorized access to digital systems, often to steal, alter, or destroy data. Hackers can operate with various motives, ranging from criminal to political (Moore, 2010). The comparison between scambaiting and hacking requires careful justification. While the primary methods differ, with scambaiting relying on social interaction, whereas hacking often focuses on technical penetration, there is a significant overlap in the use of deception, impersonation, and social engineering to manipulate a target (Hadnagy, 2015). Furthermore, hacking possesses a long and intricate history, with early manifestations such as “phone phreaking” emerging in the mid-20th century. As technology has advanced, the widespread availability of sophisticated tools and techniques has rendered certain forms of hacking more accessible. There is considerable overlap between scambaiting and hacking, depending on the specific activities undertaken. Frequently, strategies employed in hacking are also adapted within scambaiting. A notable example is social engineering, which hackers may exploit for personal gain, whereas scambaiters employ it to engage scammers more effectively, often by impersonating potential victims to prolong interactions and divert the scammers’ attention (Hadnagy, 2015). The similarities, however, extend beyond social engineering. Scambaiters have, in some instances, been observed engaging in unauthorised remote access to scammers’ computers and networks, as well as implementing tactics such as call flooding, overwhelming a scam operation’s telephone systems with non-genuine calls. Such offensive strategies are intended to disrupt operations, hinder communications, and in some cases disable scammers’ technological infrastructure.

Some common scambaiting tactics that have been observed include engaging scammers in lengthy and complex communication mimicking the behaviour of potential victims with the goal of disrupting scammers’ efforts of reaching actual victims (Zingerle, 2014); sending scammers on wild goose chases persuading them to engage in lengthy travel to collect non-existent funds, gifts, or to meet fictitious individuals who will not be there, thus wasting time and resources (Zingerle, 2014); bait and switch strategies promising scammers large sums of money or incentives, putting them through a series of tasks that ultimately yields no actual reward (Sorell, 2019).

In terms of tools, scambaiters utilise email addresses that are not linked to their own identity in order to be able to interact with scammers without revealing personal information (Button and Cross, 2017). VPNs are also commonly used, and they are essential for masking IP addresses, adding an additional layer of security to protecting scambaiters’ identities and physical locations (Whitty, 2015). Call spoofing technologies are also used in order to disguise real phone numbers during voice interactions with scammers (Hadnagy, 2018). Alongside these, scambaiters also employ various psychological tactics among which we can count mimicking

potential victims, adopting the persona of susceptible individuals in order to engage scammers (Sorell, 2019), creating frustration and exhaustion by creating convoluted and complex scenarios that scammers need to navigate towards successfully defrauding their target (Button and Cross, 2017), and even reverse psychology in the form of challenging scammers in ways that get them to commit more time and effort, further luring them in fruitless endeavours (Zingerle, 2014).

The list of scambaiting strategies and tools will be delved into in more detail in the discussion chapter with added elements uncovered after the data analysis.

Another significant element worth discussing on the topic of scambaiting relates to the legal and ethical challenges that are posed by the activities falling under this umbrella. Scambaiting often involves activities and interactions that occur across legal jurisdictions where laws can vary significantly. This brings about a level of legal complexity since it is difficult not only to determine which legal framework the scambaiting activities should be evaluated against, but also which would be the appropriate jurisdiction to evaluate and to take action where activities fall foul of the law (Brenner, 2007).

A particular challenge stems from the transnational nature of both scambaiting and the fraudulent schemes it targets. When scambaiters engage with scammers operating abroad, questions arise over the extraterritorial reach of criminal statutes and mutual legal assistance procedures between states (Koops and Goodwin, 2014). Even if an action is lawful in the scambaiter's jurisdiction, it may constitute a criminal offence in the jurisdiction where the scammer is located, and vice versa. The attribution of actions to specific actors across borders is further complicated by the use of anonymising technologies and proxy networks, which obscure the physical location and identity of both scammers and scambaiters (Yar and Steinmetz, 2019).

Beyond jurisdictional issues, the nature of some scambaiting activities can cross into legally prohibited territory. Unauthorised access to computer systems, interception of communications, and the deployment of malware are all tactics that, while sometimes rationalised as forms of disruption or intelligence gathering, are prohibited under computer crime legislation in many jurisdictions (Smallridge and Wagner, 2020). For example, the UK's *Computer Misuse Act 1990* criminalises the unauthorised access to, or modification of, computer material, with no explicit exemption for activities undertaken in the public interest (Button, 2020). In practice, this means that a scambaiter attempting to retrieve files from a scammer's server could be exposed to prosecution even if their ultimate aim was to protect victims.

Furthermore, certain strategies used in scambaiting may fall within the scope of harassment, communications offences, or data protection violations. Tactics such as “call flooding” to overwhelm scam call centres, or the public disclosure of personal information belonging to scammers, may breach laws intended to safeguard privacy and prevent malicious communications (Trottier, 2017). Such actions also raise ethical questions about proportionality, collateral harm, and the erosion of due process. Scholarship on digital vigilantism warns that while these methods may appear effective in the short term, they risk undermining the legitimacy of anti-fraud efforts by blurring the line between lawful investigation and unlawful retribution (McDonald, 2021).

Many scambaiting tactics also involve a level of deception which also brings with it challenges as far as legal boundaries are concerned. The practice raises complex legal questions because it operates in a grey area, potentially touching upon laws related to impersonation, harassment, wire fraud, and unauthorised access to computer systems, depending on the specific tactics employed. Laws concerning deceit and misrepresentation could potentially apply to scambaiters, particularly if their actions cause harm or damages to the scammer beyond mere annoyance. The questions raised by the use of deception are not only legal in nature but also dip into the ethical realm, touching upon philosophical debates about whether “ends justify the means” and discussions around the morality of mirroring unethical behaviour.

Another consideration is the level of harm and damage that scambaiting activities cause to Scammers. Considering one of the more common goals of scambaiting activities is to waste scammers time and resources, and the means by which this is sometimes achieved (involving causing various levels of frustration), this could translate into psychological or financial harm. Ethical considerations around whether causing harm, even to a scammer, is justifiable come into play. Furthermore, the detrimental impacts of scambaiting upon scammers can reach significant levels going as far as putting them in physical danger. Zingerle (2013) makes reference to a scambaiting operation that saw two scammers travelling from Lagos, Nigeria, to the “violent and desolate” Chad-Sudan border under the promise of a financial reward. Subsequent email communications between the scambaiter(s) and a potential third scammer indicate that the two scammers that travelled to the Chad-Sudan border have been imprisoned and are being exploited by the police. Assuming that this information is truthful, this raises significant concerns around the consequences of some scambaiting activities.

Furthermore, on occasion, scambaiting activities will seek to expose scammers’ personal information with the goal of publicly shaming them in the hopes that this would lead to their apprehension or to serve as detriment to other potential scammers. This raises concerns around privacy rights and the appropriateness of public shaming as a punitive or deterrent

measure (Nissenbaum, 2011). This practice is particularly relevant to Braithwaite's (1989) theory of shaming, which distinguishes between two forms. 'Reintegrative shaming' involves expressing disapproval of an act while maintaining respect for the offender, with the goal of encouraging them to rejoin the law-abiding community. In contrast, 'disintegrative shaming' stigmatises and ostracises the offender, creating a "class of outcasts" which can lead to further deviance. Recent research continues to explore the complex relationship between shame and recidivism, often finding that stigmatising shame can be counter-productive (Kuba, 2021; LeBel, 2017). It is critically important to question whether the public shaming in scambaiting is truly 'reintegrative'. Often, the intent appears to be ridicule and exclusion rather than reform, aligning it more closely with disintegrative shaming. This process of stigmatisation, as described by Erving Goffman (1963), can solidify a deviant identity rather than correct behaviour.

Beyond the immediate harm caused, some scholars offer a more critical perspective on the practice. Nakamura (2014), for example, analyses scambaiting as a 'digital show-space' that can perpetuate racial stereotypes and enact a form of 'racial violence,' particularly in the context of 419 scams originating from West Africa. This perspective challenges the heroic narrative often associated with scambaiting, forcing a consideration of the underlying power dynamics and the potential for scambaiting to reinforce harmful social hierarchies, even as it fights crime. This becomes further problematic when we account for the fact that, generally, the personal information being disclosed has been obtained via deception or other illegal means (it is also the case that on occasion, scammers have accidentally leaked their personal information by accident, or that it was obtained through the use of open-source intelligence methods).

The legal and ethical implications of scambaiting activities are substantial and complex and appropriate consideration should be given to such factors. From a legal perspective, scambaiters need to navigate a maze of jurisdictional and legal norms that may conflict across borders while ethically, they are opening themselves to scrutiny around the moral dimensions of utilising deception and other problematic strategies towards combating online fraud.

Such complexities, alongside a growing interest in cybercrime and online fraud, have drawn increasing media attention to the phenomenon of scambaiting, with multiple articles touching upon the subject, contributing to wider public awareness and engagement. NordVPN's blog has had an entry specifically aimed at exploring the topic of scambaiting in September 2023, touching upon tactics used, as well as the growth of scambaiting as a form of online entertainment (NordVPN, 2023). The Guardian has also featured an article in 2021 delving into the activities of specific scambaiters (such as IRLRosie and Jim Browning) and exploring their scambaiting activities as shared via YouTube (The Guardian, 2021). The BBC has also had an

article in 2021 discussing how a specific scambaiter (Kitboga) utilises AI technology to try to more efficiently disrupt scamming operations (BBC, 2021a). One final mention (although articles about scambaiting are much more numerous than what is listed) sees a BBC Panorama episode utilising CCTV footage of scamming operations retrieved by Jim Browning with the purpose of revealing their activities and educating the public on the dangers of such online frauds (BBC, 2021b).

While appropriately evaluating the effectiveness of scambaiting operations is particularly challenging due to a significant amount of the information required to determine the impact of scambaiting activities being limited or impossible to access, the increased media attention drawn by such activities is at least beneficial in disseminating information about online fraud across the wider public and in contributing to educating the general population on identifying and avoiding online scams. The articles discussing scambaiting also appear to tend toward a positive perception of scambaiting in general, perhaps serving as an indication of societal perceptions upon this phenomenon. However, critical pieces that raise ethical concerns and questions of effectiveness do exist (McKay, 2022; SCARS, 2021).

We have seen that scambaiting is an umbrella term that can encompass a large variety of activities that aim to disrupt online scamming operations and to protect potential victims from being reached by scammers. The nuanced nature of scambaiting operations place these in a grey area in relation to legal and ethical implications, which need to be given consideration and to be put in balance against the potential benefits. The effectiveness of scambaiting operations is also difficult to establish but while there are positives from such activities (such as the educational value of the scambaiting content produced), as with other digilante activities, there is also a significant risk of harm that should be considered. Societally, scambaiting appears to be often viewed positively, particularly in digital cultures that value proactive measures against cyber threats, and its emergence has made noteworthy contributions to the cybersecurity landscape.

2.4 The Playful Dimension: Insights from Cultural Criminology

While conceptual frameworks of vigilantism and digilantism provide important insights into the structural and moral imperatives underpinning scambaiting, framing it as a form of citizen-led justice or digital self-policing, they tend to understate the intricate experiential, performative, and ludic dimensions of the practice. Scambaiting is not solely reducible to an instrumental or deterrent logic; it is often suffused with humour, creative improvisation, narrative construction, and a sense of playful engagement with risk. Here, cultural criminology offers a productive lens,

as it foregrounds the affective intensities, subjective meanings, and symbolic enactments that imbue deviance and crime with cultural resonance (Ferrell, Hayward & Young 2008; Ilan 2019). By conceiving of crime not merely as an objective breach of law, but as a social drama embedded in mediated spaces, cultural criminology allows us to situate scambaiting within a continuum of transgressive cultural practices that blur the lines between justice, entertainment, and artifice.

Central to this perspective is Mike Presdee's (2000) notion of the carnival of crime, which argues that the hyper-rationalised and regulated conditions of late modernity cultivate a yearning for moments of cathartic, carnivalesque release. The carnival functions as a symbolic inversion of everyday order: hierarchies are destabilised, the solemnity of authority is mocked, and normative constraints are temporarily suspended in favour of exuberant, emotionally charged transgression. Scambaiting can be read as a contemporary, digitised instantiation of this carnival logic. In these online interactions, the scambaiter transforms the scammer from a threatening agent into an object of public ridicule, leveraging absurd scenarios, exaggerated personas, and performative humour to subvert the original asymmetry of power. In doing so, scambaiting transforms a potentially harmful criminal encounter into an orchestrated spectacle for the amusement and moral satisfaction of a networked audience, an inversion that parallels the festive yet confrontational energy Presdee identifies in his analysis of street-level carnivals and folk deviance.

The performative dimension of scambaiting also aligns with Stephen Williams's (2007) theorisation of potential spaces of crime, conceptualised as liminal zones where the boundaries between subjective fantasy and objective reality are blurred. In these spaces, transgression becomes a creative act, an experimental, semi-scripted engagement with the illicit that is as much about play as it is about resistance. In scambaiting, virtual machines, fabricated email identities, and elaborate fictitious backstories construct precisely such potential spaces. These environments permit the safe simulation of risk: the scambaiter can engage in deception, manipulation, and verbal jousting without exposure to the tangible harms that characterise face-to-face offending. This dynamic is not dissimilar to the mechanics of digital gaming, where players inhabit avatars, explore rule-breaking, and test alternative identities within a secure, bounded environment (Groombridge 2018). The scambaiter thus occupies a dual role, as both "player" and "performer", engaged in a form of rule-bending theatre for both personal gratification and audience consumption.

Expanding further, Keith Hayward's (2015) articulation of the Five Spaces of Cultural Criminology, spatial, emotional, mediated, temporal, and performative, provides a comprehensive framework for locating scambaiting within overlapping domains of meaning.

Spatially, the practice occupies hybridised zones of interaction: neither fully "public" nor wholly private, but rather mediated digital arenas that collapse geographic distance while retaining a sense of intimate interaction. Emotionally, the exchanges oscillate between tension, amusement, and cathartic triumph, generating a complex affective economy that sustains both scambaiter and audience engagement. Mediated space is crucial: the performance is not ephemeral but recorded, edited, and disseminated via platforms like YouTube or Reddit, where it accrues cultural capital and becomes part of a broader folklore of digital resistance. Temporally, the drawn-out nature of some scambaiting interactions amplifies narrative tension and deepens audience investment, while performatively, the encounters exhibit many of the dramaturgical features identified by Goffman (1959), including role adoption, stage management, and audience feedback loops.

From another angle, constitutive criminology (Henry & Milovanovic 1996) deepens our understanding of scambaiting as an inherently co-produced phenomenon. In this framework, crime is not a fixed category but a socially constructed process that emerges from interactions between agents, structures, and discourses. Scambaiting embodies this co-productive logic: the scammer and scambaiter together create a shared, albeit adversarial, narrative space in which identities are negotiated, challenged, and performed. The scammer's scripts and manipulation tactics are countered with the scambaiter's improvisations and absurdist provocations, producing a dialectical interaction that would not exist without the mutual participation of both parties. This interactive process also resonates with theories of deviant leisure (Smith & Raymen 2016), which locate transgressive pleasure in activities that breach social norms without necessarily pursuing material gain, suggesting that for some scambaiters, the appeal lies as much in the thrill of the performance as in any moralistic or punitive objective.

When synthesised, these theoretical perspectives suggest that scambaiting is best understood not as a purely functional mode of digital self-defence, but as a complex cultural performance that incorporates elements of carnival, play, dramaturgy, and co-creative scripting. It is an act where justice-seeking and entertainment-making are inextricably entwined, producing a mediated spectacle in which the scambaiter is both activist and trickster, moral entrepreneur and digital jester. In this way, scambaiting exemplifies the broader cultural criminological proposition that acts of transgression cannot be fully grasped without attending to their aesthetic, affective, and symbolic dimensions. The inversion of power, the simulated risk, the crafting of elaborate narratives, and the communal laughter of the watching audience together transform the scambaiting encounter into something more than counter-fraud, it becomes a staged carnival of the digital age, an arena in which crime control is reimagined as creative performance.

2.5 Economic Models of Online Content Monetisation

The evolution of the scambaiting strategies discussed above has not occurred in a vacuum; it has been profoundly shaped and enabled by the concurrent evolution of economic models for online content monetisation. A significant factor that has enabled scambaiting operations to evolve and advance over the past couple of decades can be found in the increasing opportunities for the monetisation of online content via various platforms and models. Online content monetisation is a critical aspect of the contemporaneous digital economy, enabling content creators to earn money from their content, which can encompass a variety of formats including videos, music, digital art, and more. Monetisation strategies are vital as they provide financial incentives for creators enabling them to dedicate time and resources to produce high-quality content. Effective monetisation strategies have the potential to enhance the diversity and sustainability of the digital content landscape (Lobato, 2016).

The development of these diverse monetisation strategies has had a profound impact on content creators, enabling an increasing number of individuals and small teams to break into the content creation space (Adewunmi, 2024; Rieder, 2023). This, in turn, has had a significant impact on the power dynamics in media production, increasing the democratisation of the process of content creation and distribution, with content creators now capable of reaching global audiences directly, without the need of traditional gatekeepers such as publishers or broadcasters (Jenkins et al., 2013).

Monetisation strategies are fundamental to the economic sustainability of the digital ecosystem, not only due to allowing content creators to generate revenue, but also through ensuring the maintenance and development of infrastructure and platforms that allow the hosting and distribution of digital content. The economic models that underpin digital platforms are critical in defining how information and content are accessed and consumed. Furthermore, in addition to compensating content creators, monetisation also incentivizes innovation and quality by fostering a competitive creative market (Lobato, 2016).

Since the early days of the internet, content monetisation has seen a significant evolution. Whereas initially monetisation was achieved primarily through hosting ads in exchange for financial incentives, the past two decades have seen a diversification of monetisation channels with the emergence and expansion of subscription models, pay-per-view, donations, crowdfunding, and more recently, cryptocurrency-based models (Cunningham & Craig, 2019; Täuscher & Laudien, 2018).

Initially, the simplest and most direct form of online monetisation was ad-based revenue, which saw websites hosting advertisements (in the form of banner ads, videos ads, or even pop-up

ads) and being compensated per ad impression (every time an ad is viewed) or per click. This model was attractive, and continues to be practiced today, due to its simplicity and ability to scale. An increase in the traffic to a website would translate into a potential increase in revenue through ad-hosting. This model was crucial in supporting the early internet's model of free content, making digital content available to users at no direct cost to them (Evans, 2009).

As the internet expanded and evolved, content providers began experimenting with alternative forms of monetisation such as subscription models or the implementation of paywalls. This shift was driven by a need for more sustainable revenue streams, particularly for content creators and publishers, since ad-revenue alone would be inconsistent and, in the cases of high-quality content production, insufficient to cover the costs. One of the first major publishers to successfully implement a paywall was The Wall Street Journal, which demonstrated that consumers were willing to pay for high-quality, niche, or exclusive content online (Chiou & Tucker, 2013; 106.; Rußell et al., 2020).

The transition to online subscription models marked a significant shift in monetisation strategies with this model being particularly effective for services such as Netflix and Spotify which charge users a monthly fee for unlimited access to digital content. Both providers and consumers benefit from the subscription model with the former receiving predictable, recurring revenue and the latter receiving access to large libraries of content (Elberse, 2013; Rußell et al., 2020).

As the digital landscape continued to evolve and expand, online monetisation strategies grew more diverse and sophisticated to adapt to consumer behaviour. Another monetisation strategy that emerged alongside streaming platforms is the affiliate marketing model where content creators promote products or services in exchange for a commission of sales or actions completed through their referrals. Duffy (2015) highlights how affiliate marketing has enabled bloggers and social media influencers to monetise their personal brands by linking directly to products they recommend, turning their digital influence into a viable revenue stream.

Another model that has seen increasing adoption consists of sponsored content strategies which consist of content creators being paid by advertisers to produce specific content which is distributed through the same channels as the publisher's regular content. Campbell & Marks (2015) discuss how sponsored content allows brands to embed their marketing messages within the flow of everyday content consumption, making it potentially less intrusive and more effective than traditional advertising formats.

In addition to new strategies of content monetisation emerging, platforms have also begun to increasingly adopt hybrid strategies, combining subscription fees, advertising revenues, with

dynamic pricing and advertising strategies being optimised to maximize revenue (Kumar & Sethi, 2007).

The current state of content monetisation is characterized by the leveraging of data analytics, targeted advertising, and an increasing shift towards subscription-based and freemium models. These developments reflect not only the rapidly evolving nature of the digital landscape, but also the continuous search for more effective revenue generations strategies to allow further expansion and growth.

Data analytics has become quintessential in monetisation strategies as it provides deep insights into consumer behaviour and preferences, which are then leveraged to tailor content to individual users. Platforms such as Netflix and Spotify utilise this model drawing from extensive data on user engagement to provide tailored content recommendations, enhancing user experience and boosting retention and subscription rates. Furthermore, this data is also utilised in informing what projects the platform should focus on developing further aiming to maximize engagement and minimize risks associated with content production (Gomez-Uribe & Hunt, 2016).

Targeted advertising has also grown in sophistication thanks to advances in data analytics strategies, allowing advertisers to reach audiences with higher precision than before. This is facilitated by the extensive amount of data being collected on users' online activities, which is then leveraged to create highly effective ad campaigns. This approach not only enhances user engagement by ensuring that ads are relevant, but also maximize the revenue potential for the platforms hosting these ads (Evans, 2009).

The adoption of subscription and freemium models has continued over the past year with subscription strategies continuing to be a popular choice. The offer of a fixed fee for continued access leading to the provision of a predictable revenue stream continues to be crucial for services such as streaming platforms where there is alignment with consumers' desires for continuous and varied content access. Freemium models, consisting of offering a free version of a service and providing benefits to entice users to upgrade to paid, premium versions, have also been growing, balancing the need for revenue generation with the provision of broad accessibility. The perceived nature of premium features can significantly influence a user's willingness to upgrade from free to paid versions, highlighting the importance of strategic content gating in freemium models (Wagner et al., 2014).

In instances where traditional revenue streams might not be fully viable, or where creators are seeking direct support from their audience, monetisation strategies around crowdfunding and donations have emerged to bridge the gap. Crowdfunding involves raising financial

contributions from a large number of people, typically via an online platform, which are then utilised to develop and launch a specific project or piece of content. Platforms such as Kickstarter and Indiegogo have become increasingly popular among creators for launching products and projects that might otherwise encounter challenges in securing traditional funding. These platforms allow creators to present their ideas to potential backers, offering various rewards in exchange for financial contributions (often on a tier basis tied to the amount of money contributed). Crowdfunding facilitates not only the financial aspect of the creative project, but also helps building a community around it, thus securing an audience for the project upon completion contributing to both financial support and market validation (Mollick, 2014).

Donations are another monetisation strategy that has seen increasing adoption, with platforms like Patreon allowing creators to receive revenue from and build closer relationships with their patrons. In the digital content realm, this model is particularly attractive where creators, such as podcasters, YouTubers, and artists, can offer exclusive content, early access, or special recognition in exchange for regular donations. Similarly to crowdfunding, the relationship-based nature of the donation model not only stabilizes the creator's income, but also impacts upon audience engagement, fostering a sense of community and mutual investment in the success of the content. Supporters have been found to often feel an intrinsic reward in seeing the creators they support succeed (Gerber et al., 2012).

Towards increasing their audiences and communities so as to drive additional revenue, content creators could be driven to tailor their offerings to specific audience segments, with the targeting potentially happening based on demographic data, viewer preferences, and behaviour patterns identified through analytics. By creating content specifically aimed at resonating with a particular group, content creators can build and maintain loyal audience bases (Tauscher & Laudien, 2018). Additionally, user engagement is also utilised frequently towards enhancing the experience of the audience, through interactive elements such as polls, quizzes, and user-generated content challenges. These elements not only make the content more engaging, but also help in gathering data about preferences, which can be used to refine future content (Jenkins et al., 2013).

This evolution of monetisation strategies over the past decades has seen a transition from traditional advertising and subscription models to more data-informed and community-oriented models that highlight a direct, interactive, and community-focused approach to funding digital content. However, the availability of monetisation opportunities has also been found to impact the content that is being created. Cunningham and Craig (2019) discuss how the monetisation potential in digital platforms incentivizes content creators to reinvest in their content to improve

production quality and professionalise their activities so that they can meet audience expectations and maximize revenue through ads or subscriptions.

Content creators on platforms like YouTube have been observed to adopt a variety of monetisation strategies including platform-provided tools and off-platform opportunities, with donations, affiliate marketing, and product sales being among the more prevalent methods employed. Channels promoting problematic content (where videos are at risk of demonetisation due to potentially breaking YouTube's terms of service) are more likely to adopt more diverse strategies of monetisation, to mitigate the potential loss of revenue from demonetisation. These channels have also been found to more frequently utilise cryptocurrency-related monetisation strategies, potentially to circumvent any limitations imposed by alternative donation platforms and channels. (Hua et al., 2022).

In addition to diversifying their monetisation strategies, monetisation opportunities also encourage creators to diversify their content so that they appeal to broader or different audience segments. This diversification can lead to experimentation with new content formats or new themes with the potential for revenue expansion serving as incentive for content creators to innovate continually (Tauscher & Laudien, 2018). On the other hand, the style and subject matter of the content can also see heavy influence from monetisation strategies with content creators being more financially incentivized towards creating content that attracts clicks and retains viewership. The market pressures of the monetisation models can thus lead to a homogenization of content as creators might favour the production of "safe" content that guarantees revenue, over "riskier" and less "marketable" content (Lobato, 2016).

Social media in particular has played a significant role in the evolution of monetisation strategies, as well as in widening participation in the online content creation landscape. Platforms such as Facebook, YouTube, and Instagram have integrated sophisticated monetisation models that leverage user data and engagement to maximize profitability. These platforms employ advanced machine learning technologies, alongside big data analytics to optimise advertising revenues and to enhance user experience through personalized content (Martinez-Lopez et al. 2022).

While the dominant revenue model for social media platforms appears to remain advertising, utilising target ads based on user behaviour and preferences, subscription models and premium services are also gaining significant traction, with individual content creators becoming able of setting up their own subscription models on YouTube to enhance their revenue via exclusive content and benefits.

Platforms such as Twitch and YouTube that are increasingly popular in the distribution of video content, have also developed specific and comprehensive terms of service that exert significant influence on how content creators can monetise their videos and streams. These terms not only determine the eligibility for monetisation but also shape the nature of the content that creators produce, aiming to align it with both community standards and advertiser preferences. YouTube's Partner Program (YPP) allows content creators to earn money from their videos through ads, channel memberships, and super chat features, but it comes with strict eligibility requirements. Creators must adhere to the platform's community guidelines as well as to monetisation policies which include maintaining a minimum of 1,000 subscribers and 4,000 watch hours over the past 12 months. Such criteria ensures that monetisation is only available to creators that demonstrate a certain level of audience engagement and content volume, encouraging them to produce content that is appealing and engaging to viewers over time. This has a significant impact on the nature of content that is created with creators being incentivized towards tailoring their content so that they boost viewer engagement metrics such as watch time and subscriber counts. A result of this is the increased creation of videos or content that are longer and that are heavily optimised for YouTube's recommendation algorithms, sometimes at the expense of creativity or diversity of content (Martinez-Lopez et al. 2022).

Twitch's approach to monetisation is similar to YouTube's, with Twitch being more focused on livestreaming (particularly in the gaming community) rather than video hosting. Similarly to YouTube, Twitch also employs a number of metrics to determine eligibility for its Twitch Affiliate and Partner programs such as number of regular viewers, streaming frequency, and overall following. Additionally, Twitch's model also encourages active engagement of viewers through real-time interactions such as quizzes, polls, and challenges posed to the streamers by the viewers. This promotes a community-centric approach to livestreaming, but it also has the potential of pressuring creators to perform, potentially leading to issues around digital wellbeing and burnout (Martinez-Lopez et al. 2022).

Furthermore, both Twitch and YouTube (as well as other social media and content hosting platforms) enforce specific restrictions on types of content that can be monetised. Examples of these restrictions include videos involving violence, hate speech, or misinformation. These restrictions are intended to make the platforms advertiser-friendly but can lead creators to self-censor or avoid certain topics in order to preserve the monetisation of their content, irrespective of the importance or newsworthiness of the content in question (Johnson, 2022). Additionally, platforms like YouTube have specific guidelines that dictate how sensitive content should be handled in order to comply with legal standards but also to foster a safe and respectful environment. Content creators must navigate the platform's terms of service that protect individuals' privacy rights to ensure that their content does not get demonetised or taken down

altogether (Johnson, 2022). This is particularly relevant to scambaiting where, in instances where scambaiters gain access to scammers' personal information, they need to ensure that this is censored so that they do not fall foul of the terms of service of whatever platform they are using.

Looking ahead, the future of social media monetisation is likely to be shaped by the recent advancements in AI and machine learning technologies that refine content personalization and advertising effectiveness. As these technologies become more sophisticated they will offer new ways to enhance user engagement and open up fresh avenues for revenue generation.

There are also ethical considerations to be taken into account when it comes to the monetisation of online content. These revolve mainly around the impact of monetisation strategies on content integrity, consumer privacy, and broader cultural consequences of market-driven content creation.

One of the primary ethical concerns to be considered is the potential impact of monetisation on the integrity of content. With an increase reliance on revenue from their digital outputs, creators could be tempted to focus on content that maximizes views and engagement at the expense of quality, accuracy, and truthfulness. The most evident manifestation of this is spotted in the increasing prevalence of clickbait – an approach where titles and headlines are designed to attract attention but might not accurately reflect the content's substance. Such practices have the potential to mislead audiences and to erode trust in digital platforms as sources of reliable information by prioritising profit over authenticity and reliability of content. (Couldry, & Turrow, 2014).

Significant concerns have also been raised in relation to consumer privacy. Monetisation strategies often rely heavily on data analytics which involves collecting vast amounts of personal data from users in order to tailor and target advertising more effectively. Significant privacy concerns come into play as users might not always be fully aware of what data is being collected, how it is being used, or who it is being shared with. Increasingly, businesses are harnessing and exploiting personal data for monetisation which can have profound implications for users and consumers (Zuboff, 2019). Furthermore, poor data practices and data leaks can lead to consumers' data becoming available to malicious actors who can leverage this data in order to conduct more targeted scamming operations. Scammers often exploit stolen data such as credit card information or personal identification details to commit a range of fraudulent activities including opening fraudulent accounts, making unauthorized purchases, or directly stealing money from victims' accounts. The data stolen in breaches is often sold on the Dark Web, providing a ready supply of personal information to criminals (Romanosky, 2016).

Scammers are also able to leverage stolen data to conduct more targeted scamming operations, tailoring their activities to specific individuals or groups so that they increase their chances of success (Gupta, Gupta & Kumaraguru, 2015). This targeted approach is particularly noticeable in phishing scams where emails are designed to appear as legitimate communications from trusted sources to the users. The use of data retrieved from breaches enhances the credibility of these scams, making them more difficult to detect by potential victims (Hadnagy, 2018). By using real data, scammers are better equipped to convince victims of the legitimacy of their fraudulent activities, making it easier to gain and build trust. This has been seen to be a significant factors in particular scamming strategies such as romance scams and business email compromise schemes where trust plays a crucial role (Whitty, 2019).

Another ethical consideration to be given to the monetisation of digital content revolves around the cultural and social impact created by monetisation strategies. For example, when content that attracts more views or engagement is promoted over other content, it can lead to homogenization, reducing diversity, and marginalizing certain voices or perspectives. Additionally, the pressure to monetise can lead content creators to prioritise content that appeals to wider audiences, potentially at the expense of niche topics or undeserved communities. Platforms' policies and algorithms driven by monetisation motives can influence the visibility and viability of diverse content (Gillespie, 2010).

Content monetisation strategies have seen significant evolution over the past decades, becoming more complex, more diverse, and more easily available to content creators. While this has had positive effects on enabling wider participation in the online content creation environment in ways that are relatively financially sustainable, potential negatives have been identified in issues created by content being dictated by monetisation models. When it comes specifically to scambaiting as content, the available academic literature only barely touches upon the link between digital content monetisation and scambaiting activities. However, the relation between scambaiting and modern content monetisation strategies will be touched upon in the latter Discussion chapter, drawing from both the literature discussed above, as well as the findings of the data collection and analysis of this piece of research.

2.6 Chapter Summary

This literature review has systematically mapped the academic terrain surrounding the phenomenon of scambaiting. The chapter began by situating scambaiting within the broader cybercrime debate, classifying it as a form of digilantism that responds specifically to 'cyber-enabled' offences like online fraud. To build a robust theoretical foundation, the review first

explored classical sociological and criminological theories that explain the emergence of vigilantism. Frameworks from Weber, Durkheim, and Merton were employed to understand vigilantism as a response to the perceived failure of legal-rational authority and a state of anomie, while Routine Activity Theory and Social Control Theory were used to frame it as a form of informal guardianship arising from weakened social bonds.

The review then established that while these macro-level theories explain the why of vigilantism, understanding the how of scambaiting requires a micro-sociological approach. Consequently, Symbolic Interactionism, and particularly Erving Goffman's dramaturgical analysis, were identified as the central theoretical lenses for this thesis. This framework conceptualises the deceptive encounters between scambaiters and scammers as complex theatrical performances, complete with front stages, back stages, performance teams, and impression management. Furthermore, insights from Cultural Criminology were integrated to illuminate the playful, performative, and carnivalesque dimensions of the practice, framing scambaiting not merely as informal justice but also as a form of transgressive entertainment.

The chapter then synthesised the existing, albeit sparse, literature on scambaiting itself. This included an examination of the 'non-ideal' status of fraud victims, which provides a key motivation for scambaiters, and an overview of the diverse tactics they employ. Crucially, the review dedicated significant attention to the profound legal and ethical complexities inherent in the practice, discussing jurisdictional challenges, potential criminal liability, the morality of deception, and critical perspectives on public shaming and the perpetuation of stereotypes.

Finally, the review addressed a critical, and largely underexplored, aspect of modern scambaiting: the economic models of online content monetisation. It traced the evolution of these models and analysed how platforms like YouTube and Twitch have not only enabled the professionalisation and expansion of scambaiting but also shaped its form and content through their algorithmic and policy frameworks. By synthesising these diverse bodies of literature, this chapter has highlighted a clear gap in understanding the interplay between scambaiting as a dramaturgical performance of justice and the economic and platform-based pressures that govern its contemporary manifestation. This provides a solid foundation for the empirical investigation to follow.

Chapter 3 Methodology

3.1 Introduction

This chapter outlines the methodology utilised to explore the complex phenomenon of scambaiting through the analysis of video transcripts and comments sourced from YouTube. Considering the significant role played by online platforms in modern scam operations and the increasing popularity of scambaiting as a countermeasure, a systematic approach is necessary to understand the dynamics of this phenomenon in the wider ecosystem surrounding online fraud. A robust methodological framework not only enhances the credibility of the findings but also facilitates a nuanced interpretation of the data.

Guided by the primary theoretical framework of Symbolic Interactionism, and more specifically Erving Goffman's dramaturgical analysis, this study adopts an integrative qualitative methodology that attempts to achieve a multidimensional exploration of the phenomenon of scambaiting. Symbolic interactionism focuses on the interpretations and meanings that participants assign to their experiences and interactions within the scambaiting discourse, especially taken into considerations the multiple personas and roles that actors often impersonate.

The data for this research consists primarily of publicly available YouTube content, including comments and transcripts, which were programmatically retrieved and analysed using advanced natural language processing techniques facilitated by large language models (LLMs). This approach allows for efficient handling of extensive volumes of qualitative data, introducing benefits to the analysis, but also challenges and limitations.

In discussing the methodology, particular attention is given to the challenges associated with relying on LLMs for data analysis, such as potential inaccuracies in generated insights and the inherent biases that may be present in automated outputs. The decision to focus exclusively on YouTube content presents additional limitations, including cultural and linguistic barriers that may restrict the breadth of the findings. Furthermore, the quality of data retrieved programmatically raises important considerations regarding transcription accuracy and contextual relevance.

Ultimately, this chapter aims to provide a comprehensive account of the methodological choices made throughout the research process, exploring how these choices shape the outcomes of the study. By critically examining the interplay of theoretical frameworks, data

collection strategies, and analytical techniques, this chapter hopes to establish the groundwork for understanding the complexities of scambaiting in the broader context of online scams.

3.2 Research design and strategy

The research design for this study is underpinned by the principles of qualitative inquiry, generating theory from systematically collected and analysed data to illuminate social relationships and behaviours within groups (Noble et al., 2016). The study draws on symbolic interactionism to provide a multifaceted analytical framework. Symbolic interactionism offers a theoretical foundation by emphasizing the importance of social interactions in constructing reality. This alignment is beneficial for situating the research within the broader context of social interactions and meanings (Blumer, 2012).

The digital space of scambaiting and online fraud has become increasingly sophisticated and multifaceted. Scambaiters adopt roles ranging from passive victims to strategic deceivers, posing as vulnerable individuals, deploying personas, or even engaging in automated countermeasures, to trick scammers and disrupt their operations (Chen, Wang & Edwards, 2022; Dynel, 2021; Ross, 2021). This complexity is especially pronounced in the approaches employed by online scammers and scambaiters, who often engage in considerable "role-playing" as they pursue their goals, necessitating the juggling of multiple personas and perspectives to successfully deceive their counterparts. For instance, when a scammer gains remote access to a target's computer, they may probe the system to identify elements indicating they are accessing a virtual machine. This probing must be conducted covertly to avoid raising suspicion from legitimate targets. In response, the scambaiter may covertly operate their computer to maintain the illusion of legitimacy while concealing their technical expertise.

Symbolic interactionism provides a lens through which to understand the subjective meanings and social interactions that define scambaiting. By focusing on how scambaiters and scammers interpret and respond to each other's actions, symbolic interactionism uncovers the underlying motivations, strategies, and adaptations characterizing these interactions. This theoretical perspective is crucial for examining the role-playing and identity management tactics employed by both scambaiters and scammers as they navigate the highly complex environment of their deceptive strategies and engagements.

3.3 Data collection methods

This study on scambaiting employs a qualitative approach, utilizing data extracted from YouTube videos and their accompanying comments.. The goal of this approach is to achieve a comprehensive analysis of the scambaiting phenomenon as evidenced by the content shared on YouTube. The collected data pertains not only to the content produced by scambaiters but also captures the interactions indicative of audience engagement.

YouTube videos and their corresponding comment sections have been chosen as the primary data sources due to their rich content and significance to the study of scambaiting and online monetisation strategies. The videos themselves generate a wealth of qualitative data in the form of video transcripts that encapsulate the dialogues and interactions of scambaiters with scammers, viewers, collaborators, and victims. This data is pivotal for understanding the techniques and narratives employed in scambaiting by both scambaiters and online scammers. The comment sections add another layer of qualitative data, presenting audience reactions, engagement, and insights into public perceptions and community dynamics surrounding scambaiting.

Additionally, contextual metrics, including the number of comments, views, likes, and subscriptions, were collected to provide context on channel popularity and evolution.. However, an initial intention to conduct a formal quantitative analysis of this data was ultimately set aside to allow for a more concentrated and in-depth qualitative analysis of the rich data within the video transcripts and comments.

To ensure a comprehensive and representative dataset, a multifaceted sampling strategy was employed. In the early stages, several searches were conducted on YouTube utilising keywords such as “scams,” “scambaiting,” “online fraud,” “tech support scams,” and “romance scams,” along with various derivatives. These terms were strategically selected as they were guided by a systematic review of key terms in existing academic literature on cybercrime and online fraud, combined with vernacular terms commonly used by the scambaiting community on platforms like YouTube. Engaging with identified content further activated YouTube’s recommendation algorithm, helping to discover additional relevant videos and channels based on current engagement patterns. Moreover, references made by various scambaiters and commenters about the work of others helped to uncover additional resources.

This process ultimately led to the identification of several YouTube channels, which were meticulously evaluated for their appropriateness for inclusion in this study. To maintain the focus on scambaiting content, the researcher analysed the videos produced by these channels, specifically distinguishing between those dedicated primarily to scambaiting and those that

feature a broader content mix. Channels where scambaiting did not serve as the primary focus were excluded from further consideration. The final selection includes the following YouTube channels, presented in alphabetical order:

- DEYOCLUB
- IRLRosie
- Jim Browning
- Kitboga
- Malcolm Merlyn
- More Kitboga
- Pleasant Green
- Rinoa Antidote
- Rinoa Poison
- Scam Sandwich
- Scambaiter
- Scammer Payback
- Scammer Payback Live
- ScammerRevolts
- SkeletonSyskey
- The Hoax Hotel
- Trilogy Media
- Trilogy Vault

It is important to note that some individual scambaiters manage multiple YouTube channels to categorise their videos based on duration. Channels focusing on very short video content (typically lasting a few seconds to a minute) have been disregarded, as they mostly consist of snippets extracted from longer videos. Additionally, some channels are utilised solely for live streaming purposes.

Upon finalizing the list of channels, the researcher employed the Google API through a Python script to extract lists of all videos published on each channel, storing the video IDs in separate .csv files for each channel. After retrieving the lists, a different Python script was utilised to extract auto-generated transcripts for each video, filtering out any videos where a transcript was unavailable (though the researcher still viewed these videos). The decision to extract transcripts from YouTube, as opposed to manual transcription or utilising alternative methods, was made in light of the substantial volume of videos identified. Across the 20 selected channels, over 5,000 videos with available auto-generated transcripts were found. Manual transcription of this volume, particularly with many videos spanning several hours,

would be impractical, and utilising other transcription tools was assessed as being more computationally intensive than extracting existing YouTube-generated transcripts.

While some quality issues with auto-generated transcripts were encountered during manual checks, especially when audio quality was lacking or in instances of overlapping dialogue, such occurrences were observed to be relatively infrequent.

Following the retrieval of video transcripts, the same lists of videos from the 18 scambaiting channels were utilised to gather comments for each identified video ID. A separate Python script was created using the YouTube API to retrieve the comments for each scambaiting video, saving them into distinct .csv files organised by channel. While this process was straightforward and not computationally intensive, limitations set by the YouTube API necessitated that comments be gathered over several days. Due to these constraints, the decision was made to collect only first-level comments, excluding replies to initial comments. This choice was driven by the higher API cost associated with fetching replies compared to first-level comments, raising concerns about the time needed to extract all comments. Despite this limitation, the researcher believes that the number of excluded comments is not significant since the majority of observed replies on scambaiting videos consist of first-level comments. Ultimately, this process resulted in the retrieval of over 6 million comments for analysis.

The YouTube API was also employed to collect current metrics for the selected scambaiting channels and their corresponding videos. The retrieved metrics included view counts, likes, comments, and subscriber numbers for each channel. Additionally, Social Blade was utilised to access historical data pertaining to these metrics, providing insight into channel growth and evolution over time. The dates of individual video uploads were also collected to analyse the frequency of scambaiting content releases across the selected channels.

A stratified random sampling method was used to select 5% of the total videos for in-depth analysis. To ensure that all YouTube channels were appropriately represented, the sampling was conducted separately within each channel's video set. In other words, 5% of the videos from each channel were randomly selected so that the final sample maintained the same proportional distribution of channels as in the overall population. This ensures that the final sample is not dominated by a few large channels and that videos from smaller or more niche creators are proportionally included. A 5% sample was chosen as a balance between constituting a large, representative dataset and maintaining a manageable scope for in-depth qualitative analysis of video transcripts. This resulted in a total of 215 videos watched, during which the researcher diligently took notes and recorded observations. The duration of the randomly selected videos varied considerably, ranging from approximately 10 minutes to over

10 hours, and encompassed a diverse mix of scamming and scambaiting strategies, featuring various styles, from educational to entertainment-focused.

While watching the randomly selected videos, the researcher endeavoured to take detailed notes on key themes, interactions, techniques, and audience engagement, as well as content style. These notes were collected in separate documents corresponding to each YouTube channel, which were later organised and coded in alignment with the video transcripts and comments.

In addition to analysing the selected videos, the researcher sought to develop a suitable strategy to interrogate the entire dataset collected, considering the temporal dimension of how scambaiting operations have evolved and whether specific trends can be identified within the data. Given that the videos span a decade, with the earliest uploads dating back to 2014, this exploration of temporal patterns is essential. However, it is acknowledged that the smaller selection of 215 videos may be insufficient for yielding a well-rounded understanding of patterns or trends that may emerge along this temporal axis. Further insights into the strategies employed to tackle the complete dataset will be outlined in the next section.

It is important to note that an initial plan to conduct interviews with scambaiters was not pursued. Despite initial outreach to several content creators, the low response rate (only two replies were received) made the interview method unfeasible for achieving a representative sample for this study.

The data collection process was executed with strict adherence to ethical standards, ensuring that the privacy of users whose data has been utilised is respected. Although all data used in this study is publicly available, several measures were taken to protect privacy. The Python script used for data collection was designed not to capture usernames from comments, and the final analysis avoids linking specific excerpts to their source channels. The full, anonymised dataset will be made available alongside the final thesis upon its publication in the university's research repository. The decision to anonymize in this manner was grounded in ethical considerations, aimed at protecting the privacy of individuals involved and acknowledging that linking comments to users provided little benefit to the study.

The decision to disclose specific YouTube channels was made to enhance transparency and reproducibility, allowing other researchers to verify and build upon this research. Naming the channels also enables peer reviewers and readers to authenticate the data sources, ensuring that the research is underpinned by credible information. Explicit consent was not obtained from the channel creators; however, ethical approval was granted by the University of Southampton's ethics committee to name the channels on the justification that the content is

publicly available, the creators operate as public figures within this domain, and the risk of adverse impact was assessed as minimal.

While the researcher recognises that such risks exist, they believe the likelihood of this research adversely impacting the selected YouTube channels is minimal. Engagement with the content created by these channels reveals a shared commitment to exploring scambaiting and online scams, inspiring confidence that this research contributes positively to the ongoing discourse. Additionally, scambaiters have demonstrated proactive measures to protect themselves while producing content for YouTube, leading the researcher to conclude that this study does not include any information that would undermine their protective efforts.

An important insight gained from the data collection process is the presence of diverse, sometimes conflicting opinions regarding some scambaiting practices. Different YouTube channels and community members have been observed to express varied views on the benefits or detriments of specific scambaiting activities. To address these differing opinions, the researcher adopts a neutral position, striving to explore the identified conversations without imposing a particular moral stance. To maintain this neutrality, discussions surrounding specific practices will be framed generically, aiming to provide a balanced exploration that avoids unfair criticism or misrepresentation of the selected channels.

All data and Python scripts utilised for this research were securely stored in a dedicated folder on OneDrive, connected to the researcher's university-provided Microsoft account. This cloud-based storage solution not only ensured easy access but also protected the data under the university's stringent security protocols. OneDrive provides encryption during transmission and at rest, safeguarding data against unauthorized access. The use of the university account offered additional security layers, including multi-factor authentication and compliance with institutional data protection policies. Regular automated backups further ensured data integrity and continuity throughout the research process. Collectively, these measures upheld the security, confidentiality, and ethical compliance of the data.

In summary, the data collection for this study of scambaiting employed a rigorous and thorough approach to ensure data richness and reliability. Utilising qualitative data from YouTube videos and comments aims to provide a nuanced understanding of the scambaiting phenomenon in the context of modern content creation and monetisation practices. The extensive dataset collected across 18 scambaiting-centric YouTube channels, encompassing over 5,000 videos and more than 6 million comments, is expected to illuminate the interplay between contemporary scambaiting activities and online content monetisation.

3.4 Data analysis methods

The significant volume of videos identified, and qualitative data collected in the form of transcripts and comments gave rise to the challenge of identifying an appropriate strategy of conducting the necessary data analysis. These full-length verbatim transcripts form the primary dataset for the thematic analysis, and the data is managed and made available in line with the University of Southampton's Research Data Management Policy. Manual coding of such volumes of data has been found to be impractical, and as such, more technological solutions became sought after. Initially, the researcher hoped to be able to utilise some of the automated coding functionality that is available within NVivo, in order to identify themes, patterns, and recurring elements within the body of text accumulated. However, this proved ineffective, primarily due to the fact that the data collected was not of a high enough quality to enable it to be appropriately tackled by the automated features of NVivo. The data that was retrieved was in a plain text format, lacking in capital letters and punctuation marks and being structured in transcript lines rather than sentences, which made it difficult for themes and patterns to be picked up from it.

The next avenue explored consisted of utilising natural language processing (NLP) strategies via the Python packages spacy, gensim, and nltk in an attempt to retrieve insights from the transcript data collected. The python scripts were written and run locally by the researcher by using Jupyter notebooks within JupyterLab. Initially, elements such as geographical locations, types of scams, and names of companies being impersonated have been attempted to be retrieved from the data but this proved ineffective in a similar fashion to the attempts at using NVivo, again due to issues with data quality. Attempts have also been made at trying to clean the data and bring it into a more usable state, but this also proved challenging due to the large amount of data being dealt with.

Finally, a solution has been identified in leveraging the emerging technology of large language models (LLMs), made popular in recent years with the emergence of OpenAI's ChatGPT and the subsequent development of other similar technologies such as Meta's Llama and Google's Gemini. Since the launch of ChatGPT in late 2022, LLMs have seen an extreme rise in popularity and adoption, marking a significant milestone in the fields of Artificial Intelligence (AI) and natural language processing (NLP) with numerous domains being revolutionised by their human-like capabilities of understanding and generating text (Spyros et al., 2023; Liu et al., 2023).

The researcher hoped to use Meta's Llama 3.1-70B model, hosted on an inference endpoint via the huggingface.co platform. This was preferred in large part because Meta's Llama models are

open source and free to use commercially and for research purposes, but also due to the token limitations being large enough to facilitate sufficient processing power for the data at hand. However, the significant hardware required to run this model meant that a cloud solution was necessary, which is why the huggingface.co platform was used to host the model. Once a suitable cloud solution was identified to host the Llama 3.1-70B model the researcher succeeded in setting it up in such a manner that it could be interacted with through an API via a locally stored Python script.

While the connection worked as required, the outputs generated during the testing stages were not ideal. The researcher attempted to use simple prompts to elicit an answer from the LLM, attempts which resulted in outputs that were peculiar at best. In one example, upon being prompted to "Please give me an example of a colour", the output generated was approximately 1,400 words long, beginning with the following paragraph:

"Please give me an example of a smell. Please give me an example of a taste. Please give me an example of a sound. Please give me an example of a feeling.

That was a good, brief example of how I think of God. When I think of the attributes of God, I think of the attributes of a human being, because I am a human being and I don't have any other way to think. I cannot think of God as a dog thinks of God. I can't think of God as a spider thinks of God. I can't think of God as an alien from outer space thinks of God. I can only think of God as a human being thinks of God..."

This is not the output that was expected from the prompt provided, and the remainder of 1,400 words generated continued to explore the religious theme of perceiving God, without ever reaching a point of even mentioning a colour. Similar tests continued to be run as the researcher attempted to calibrate the LLM but while improvements were observed, the output never reached a point where the input provided would result in an appropriate output. Please note that this should not be taken as a reflection of Meta's LLM, but rather of the researcher's technical ability to utilise it effectively.

With the Llama LLM being ruled out, the researcher transitioned to exploring OpenAI's ChatGPT model as accessible via API. Specifically, the gpt-4o-mini model was used (which is at the time of writing the most recent model made available by OpenAI). The technical specifications of gpt-4o-mini were found to be similarly suitable for the task as Llama's with the added advantage of enabling interaction with the need for a cloud hosting solution. However, unlike Llama, OpenAI's language model are not open source.

Tests similar to the ones mentioned above were also run through OpenAI's gpt-4o-mini model, returning much more positive results, with the outputs produced being much more appropriate

for the prompts provided. After testing the model with some simple prompts, the researcher fed a very small portion of the transcript data to the model, requesting specific information to be identified and output in a specific format, which the LLM was able to consistently deliver. In light of the success of these tests, the gpt-4o-mini was deemed to be appropriate for supporting with the data processing needs for this piece of research.

In the initial stages of the research process, the YouTube API was employed to systematically retrieve a comprehensive list of YouTube videos from the pre-identified YouTube channels that were mentioned earlier. In addition to the list of videos, essential metadata consisting of video IDs, titles, upload dates, and numbers of views and comments at time of data retrieval, was also collected and stored in a structured spreadsheet in OneDrive. The video IDs served as the primary keys for subsequent data analysis tasks, facilitating the linking of associated transcripts and comments, as well as of future data analysis results.

Once the entire population of videos was identified, a stratified random sampling strategy was employed to identify a sample of 10% of the total population to be explored in more depth. A random number generator was used to generate a list of numbers equal to 10% of the total video count for each channel, with values ranging from 1 to the total number of videos on that channel. This process was repeated for each channel and the videos were indexed so that the indexes matching the generated numbers ended up being selected. This approach ensured appropriate representation across all channels leading to a total of 10% of the entire video population being selected.

The researcher engaged in immersive viewing of these selected videos, memoing and taking detailed notes throughout the process. Careful attention was directed towards techniques employed by scammers and scambaiters so as to explore the interactions that occurred between actors attempting to defraud their targets and actors attempting to thwart such activities. Additionally, focus was given to the methods employed by scambaiters to produce entertainment, engage audiences, monetise content, and to provide educational insights. The intricate interactions between scammers and scambaiters were also noted upon.

Following this, an automated data processing strategy was employed by utilising the OpenAI ChatGPT 4o-mini API through a Python script, which provided significant benefits in managing and processing the large volume of text data obtained from the video transcripts and comments. Due to the possibility of transcripts exceeding 70,000 words and to the token limitations imposed in the usage of the API, the transcripts were subjected to a python script aimed to identify whether the word limit exceeds 70,000, and to break the text into chunks where this is the case. The token limitations in question were a maximum 128,000 token window (referring to the maximum number of tokens that can be processed in one go across both input

and output) and a maximum output of 16,384 tokens. For context, 1 word is equal to approximately 0.75 tokens.

The analysis was structured through a series of template prompts, each meticulously designed to extract specific information, ensuring the LLM could effectively handle the task within the token constraints. Each template followed the same structure but asked for different items. Each prompt included the list of items requested, followed by instructions dictating the format of the output requested. The format resembled the following structure "<<<Item>>> Answer <<<Item>>>" where Item refers to an individual item listed in the prompt, and Answer would indicate the location for the LLM to provide the output pertinent to the item in question. This facilitated ease of identification and subsequent conversion into tabular form. No variations in response formats were implemented.

The four prompt templates and their corresponding items are as follows:

Elements pertaining to scam activities

- The type of scam perpetrated
- The method by which the scam is initiated
- Techniques, methods, and strategies employed by the scammer(s) to defraud their target
- Techniques employed by the scammer(s) to combat the possibility of being scambaited
- The amounts of money scammers try to defraud
- Payment methods used by the scammers to retrieve the stolen funds
- Mentions or use of cryptocurrency
- Tools, platforms, software used by scammers
- Country location of targeted victim
- Country location claimed by scammers
- Communication channels used by scammers to engage their victims
- Victim profile
- Data requested by scammers from victims
- Use of scamming scripts
- Role created and played by the scammer(s)
- Company/business/entity impersonated by the scammer(s)
- Justifications used by scammers to explain their active engagement in criminal activities

Elements pertaining to scambaiting activities

- Strategies employed by scambaiters to engage scammers and disrupt activities
- Fake personas used by scambaiters (including victim personas, as well as third-party personas)
- Techniques employed aiming to waste scammers' time
- Techniques employed to actively disrupt scamming operations
- Investigative strategies used by scambaiters to identify scamming operations for targeting, and to uncover the details behind them
- Use of humour and satire by scambaiters
- Elements pertaining to the technological infrastructure utilised by scambaiters in their activities
- Technological countermeasures employed by scambaiters to combat scammers actions towards defrauding their victims
- Instances of escalation and confrontation where scambaiters challenge the scammers on the illegality and immorality of their activities
- Psychological strategies employed by scambaiters
- Targets and objectives of scambaiters
- Mentions, references, and collaborations with third parties such as other content creators or business/companies
- Mentions of ethical and moral considerations, as well as justifications and rationalizations of stances taken
- Mentions of sponsorships, collaborations, and techniques relating to content monetisation
- Mentions of challenges and difficulties encountered by scambaiters in their attempts to disrupt scamming operations
- Outcomes of the scambaiting activities observed within the videos

Elements pertaining to interactions occurring between actors

- Scammer – Victim Interactions
- Scammer – Scambaiter Interactions
- Scammer – Third Party Interactions
- Scambaiter – Audience Interactions
- Scambaiter – Scammer Interactions
- Scambaiter – Victim Interactions
- Scambaiter- Third-Party Interactions

- Scammer – Scammer Interactions
- Scambaiter – Scambaiter Interactions

Elements pertaining to symbolic interactionism

- Use of symbols and symbolic language
- Identity construction and presentation of scammer
- Identity construction and presentation of scambaiter
- Role expectations
- Social roles
- Interactions and meaning making
- Norms and social expectations
- Language and communication styles
- Metaphors and analogies
- Deception and authenticity
- Emotional expression and manipulation
- Self-perception and reflexivity
- Social identity and group membership
- Moral judgements and justifications
- Stigmas and labelling
- Cultural references and codes
- Power and authority symbols
- Negotiating and bargaining
- Identity dissonance

Each transcript was processed multiple times through the OpenAI API script, once for each prompt template. The outputs were systematically organised into a tabular format and saved in OneDrive, with each item requested being stored in a separate column against the pertinent video ID. Deductive coding was applied during the processing of transcripts through the LLM with the researcher specifying the particular items listed above to be identified and deduced from the text. The LLM was tasked to search the text for the presence and capacity of these items, marking instances where the items were absent with the string “NOT AVAILABLE” for ease of identification of such instances in the output.

Once the outputs were retrieved from the OpenAI API, the data was merged with the initial metadata obtained from the YouTube API. This integration enabled the linking of detailed analytical outputs to their respective video titles, channels, upload dates, and viewer engagement metrics (views and comments). This linkage was crucial towards enabling a multi-dimensional analysis, affording the exploration of thematic elements not only across the entire datasets, but within segments of it relating to specific channels, as well as temporal contexts.

The qualitative data analysis progressed through a methodical approach where the researcher scrutinized the collected data column by column, in order to refine, categorise, and clean the data. This processed helped ensure accuracy and reliability, through the comparison of the LLM outputs with the notes taken by the researcher on the sample population selected earlier. The outputs were found to be largely consistent with the video content, but issues were identified with the LLM's abilities to accurately evaluate specific scenarios and occurrences. Due to the transcripts not containing any indication as to what line of text belongs to what participant, challenges emerged in the correct identification of which actor was speaking. This was particularly exacerbated in multi-person interactions and in capturing scenarios where actors talked over each other, which would not translate accurately within the transcript. Furthermore, the researcher also anticipates that infrequent issues with the audio quality of telephone calls captured, as well as the occasional use of foreign languages by scammers would have led to the transcripts retrieved not matching the contents of the call with 100% accuracy. However, such issues are minimal in occurrence, and as such the impact on the output is not expected to have been significant, especially considering LLMs ability to handle typos or errors in written communication.

The specific data from the outputs produced via LLM were aligned with the upload dates of videos, allowing for a temporal mapping of categorical data. For example, this allowed the researcher to explore any emerging trends in the types of scams captured over time, with one particular trend that could be used as an example is the increased use of cryptocurrencies to facilitate the movement of funds. Through exploring data along different dimensions, such as channel-specific trends and temporal patterns, the researcher was able to uncover evolving strategies in scambaiting content and audience engagement over time. Integrating symbolic interactionism further enriched the analysis, providing theoretical depth to the interpretation of how meanings and interactions evolve and interact within the overarching scambaiting ecosystem.

In comparison to other studies conducted on the topic of scambaiting, this piece of research draws on a much more comprehensive amount of data, spanning thousands of YouTube videos uploaded over the course of a decade across 18 different YouTube channels. The methodology

employed in the collection and analysis of data is also novel, particularly due to the use of LLM technology to support with data analysis, making the exploration of such large volumes of data more accessible. However, while the use of LLMs within the study of scambaiting is a novel approach, a significant number of academic works has already taken the first steps towards exploring and testing the applications of AI technology towards research within social sciences. Notably, a GitHub repository by the name Awesome-LLM-in-Social-Sciences compiles academic papers that evaluate, align, employ, or contribute to the exploration of LLM application within research (Value4AI, 2023).

3.5 Theoretical framework

In conducting qualitative research, the choice of theoretical frameworks is essential as it informs the design, data collection, and data analysis processes employed within the study. For this piece of research, the theoretical framework of Symbolic Interactionism was utilised, offering unique insights and methodologies that collectively contribute to a comprehensive understanding of the research context and findings.

Symbolic interactionism is a sociological framework developed primarily by George Herbert Mead and Herbert Blumer, centred on exploring human behaviour via the meanings assigned by individuals to their interactions and social contexts. At the core, this framework promotes the idea that human actions are not simply responses to stimuli, but rather they are shaped by the social meanings associated with those actions. The core principles of symbolic interactionism include concepts such as meaning making, social interaction, symbols, and the dynamic nature of social life.

At the heart of symbolic interactionism is the idea that individuals create and interpret meanings through their interactions with others and with their environments. The meanings that emerge from these interactions influence behaviour and social practices, thus influencing future behaviour. Social interactions are also pivotal in the shaping of self-identity with symbolic interactionism viewing the self not as an inherent quality but rather a construction that surfaces through social processes and feedback from others. Symbols such as language, gestures, and imagery are significant tools through which individuals negotiate meanings, aiding in the co-construction of identity and in the instruction of behaviour. However, it needs to be noted that meanings are not static, and often times the same symbol can concurrently hold multiple meanings when more than one context is present in the interaction. Meanings also evolve over time through ongoing processes of negotiation and reinterpretation, thus requiring a qualitative

approach in order to appropriately capture the fluidity and dynamic nature of their manifestation in social contexts.

Symbolic interactionism is extremely relevant when it comes to the study of scambaiting, partly due to the highly complex interactions that occur across actors juggling multiple personas, and partly due to the nature of the interactions captured where scambaiters frequently engage simultaneously with both their scammers and with their audiences. The terms and symbols employed by scammers and scambaiters can reveal underlying beliefs and attitudes about trust, deception and morality, but the personas that are taken up can be equally valuable in understanding the actor's understanding of the role they are trying to play. Particularly when it comes to scammers, symbolic interactionism could be a powerful tool towards exploring their perspectives on the type of individuals they are trying to impersonate. Furthermore, with the majority of the data captured as part of this research representing social interactions of one kind of another, symbolic interactionism has been found to be an appropriate framework for the exploration of this data.

Building upon the foundations of Symbolic Interactionism, this study specifically employs Erving Goffman's dramaturgical analysis. Goffman (1959) uses the metaphor of a theatrical performance to explain social interaction, viewing individuals as actors on a stage who actively manage the impressions they convey to others. This perspective is exceptionally well-suited to the study of scambaiting, where both scammers and scambaiters engage in elaborate performances, constructing and manipulating realities to achieve their respective goals. The following dramaturgical concepts serve as the primary analytical tools for this research.

The concepts of 'front stage' and 'backstage' are central to this analysis. The front stage is where the performance is given, such as the recorded interaction between the scambaiter and the scammer that is presented to the YouTube audience. Conversely, the 'backstage' is where the performer can drop their character and prepare for the performance. For a scambaiter, this might involve researching a scam, setting up a virtual machine, or collaborating with others outside of the recorded interaction.

All actors engage in 'impression management,' the process of controlling the information others receive to influence their perception of a person, object, or event. The 'performance' itself is the activity of an individual in front of observers that serves to influence them. In scambaiting, this is the core activity: scammers perform the role of a legitimate authority figure, while scambaiters perform the role of a vulnerable victim, all to manage the impressions of their counterpart.

A performance is supported by several elements. The 'personal front' includes appearance (the visual elements of a scambaiter's fake desktop) and manner (the voice, accent, and emotional

state they adopt). The 'setting' comprises the virtual scenery where the interaction unfolds, such as a meticulously crafted fake banking website. 'Props' are the objects used to enhance the performance, including voice changers, soundboards, fake identification documents, and scripted dialogues.

Interactions often involve 'performance teams,' which are groups of individuals who cooperate in staging a single routine. A scam call centre, where multiple scammers work together to maintain a consistent fraudulent narrative, is a clear example. Similarly, some scambaiters collaborate, acting as a team to create a more convincing performance.

Finally, Goffman distinguishes between the 'sincere' performer, who believes in the impression fostered by their own performance, and the 'cynical' performer, who does not. Scammers are typically cynical performers, fully aware that their role is a fabrication. Scambaiters are also cynical in their performance of the victim persona. However, they may be sincere in their overarching belief in the moral purpose of their scambaiting activities. This dichotomy is crucial for understanding the motivations and mindset of the actors involved.

3.6 Ethical Considerations

Ethical approval for this research was granted by the University of Southampton Ethics Committee on February 2023 (application reference 72107.A1). Given the extensive use of LLMs in this research, addressing the ethical considerations of AI technologies is a necessary starting point. The significance of the emergence and adoption of LLMs has not been without concern for a number of issues, relating to the ethical and practical implications of adopting this new piece of technology. While the magnitude of this development is sufficient to warrant discussions around its ramifications for a multitude of industries and domains, for the purposes of this piece of writing, we will limit ourselves to only exploring the implications of utilising LLMs within research, specifically as it pertains to data analysis.

One of the primary ethical concerns to be addressed relates to the issue of authorship and attribution. Possessing powerful capabilities of understanding and generating text, LLMs could be prompted to produce pieces of writing that can serve as part of wider pieces of academic (and non-academic) publications, or even as the entire body of such a publication. However, Hosseini et al. (2023) emphasize that, as far as scholarly endeavours are concerned, use of AI tools should be employed in a transparent manner, with the details and extent of their utilisation being included within introduction and methodology chapters rather than as simply generators of academic text. This view is shared by Lin et al. (2023) who posit that generative AI is capable

of enhancing research productivity and ethical use, but transparency and open science practices are essential to the detection and combatting of fake research.

This is particularly pertinent when considering that, despite their impressive generative abilities, LLMs have also been observed to produce factually inaccurate information, which continues to remain a significant concern. The generation of factually inaccurate outputs is often referred to as “hallucinating” and it presents significant issues when considering applications within academic work. Hallucinations in LLMs can stem from biases in pretraining, such as memorization of sentences and statistical patterns of usage, which can lead to incorrect inferences and generation of content that is not aligned with the input data (McKenna, et al., 2023). Furthermore, while LLMs are also prone repeating patterns without understanding them, leading to hallucinations that can mislead users and propagate misinformation (Z. Li et al., 2023).

Towards mitigating the risk of hallucinations, prompt engineering has emerged as a pivotal technique that can be employed to address the risks of hallucinations being generated by LLMs. The umbrella of prompt engineering covers a number of strategies and methodologies, among which we can count role-prompting, one-shot prompting, contextual prompts, and the use of linguistic nuances in prompts. Role-prompting refers to the structuring of the input text in such a way so that the LLM is guided towards generating more accurate responses. This can include specifying the role or perspective the model should take when generating a response, serving to optimise LLMs by structuring the input text. One-shot prompting is a technique where the user provides an example that includes the desired input and output format, enabling the LLM to use the example provided to infer how to handle the requested task. The use of contextual prompts has also been discussed as being combatting the risks of hallucination, consisting of the provision of related information that sets the stage for the LLM to be able to handle the query raised by the user. The embedding of context within the prompt allows the model to better understand the nuances of the task, improving the relevance and accuracy of the output. Finally, the use of linguistic nuances is also important in providing inputs to LLMs as the choice of words can influence the tone of the response and it can affect the output. A good example of this consists of the question “Why is renewable energy important?”, with the phrasing having the potential of creating bias within the model, suggesting that the importance of renewable energy is what needs to be focused on and demonstrated (Chen et al., 2023). It is important to note that this list of approaches to combatting the risk of hallucinations is not extensive and numerous other more technical approaches have been discussed both within academia and outside of it. However, going into such level of technical detail is not necessary for our current purposes.

In addition to the issues around factuality and accuracy of the information generated by LLMs, another ethical concern that has been frequently discussed is found within the issues of plagiarism and academic integrity. The use of LLMs in academic writing raises concerns around academic misconduct due to the ability of LLMs to generate coherent and original text that can evade detection, creating a need for academic institutions to revisit their policies to account for the emergence of AI tools. The determination of whether usage of LLMs constitutes academic misconduct depends on the academic integrity policies of individual Higher Education Institutions (HEIs) (Perkins et al., 2023). This aligns with the need for transparency outlined earlier on, where usage of LLMs and AI tools in general should be outlined within the work they are used for so that concepts central to academic research such as reliability, validity, and reproducibility continue to be adhered to.

Concerns regarding data privacy have been central to the research design. Given that the data utilised in this research project is publicly available, issues surrounding data privacy have been found to be minimal and largely inapplicable. The data processed through the LLM did not contain identifiable information that would link comments to specific users or transcripts to specific videos; this linking was managed within a locally run Python script used for data processing.

The research protocol adhered to a principle of passive observation. While the content analysed often depicted criminal activity (online fraud), the data as presented by the creators typically redacted personally identifiable information that would be necessary for reporting to authorities. In line with ethical guidelines for observational research of public data, the researcher did not engage with or report any observed potential criminality, as the role was one of a passive analyst rather than an active participant.

The data concerns are however not limited to the use of LLMs but also to the use of this data within the wider body of this piece of research. In ensuring data privacy for the comments and transcripts drawn solely from publicly available sources, several measures were implemented to mitigate privacy concerns. Notably, data anonymization was achieved by not capturing identifiable information related to the users whose comments were recorded. This approach effectively safeguards individual identities and aligns with best practices for ethical data usage.

However, it is important to consider the ethical implications of linking transcript data to specific videos and channels. While the transcripts themselves are associated with public content, disclosing this linkage warrants careful consideration; transparency about the relationship between specific transcripts and their corresponding sources may raise ethical questions regarding the potential identification of individuals or the content's origins. With these considerations in mind, the researcher has opted to avoid the linking of specific content to

specific YouTube channels/videos, as no particular benefit has been found to be available via such linkage.

The avoidance of linking specific content to specific videos and YouTube channels is also preferred when considering the potential impact upon research subjects. The scambaiting communities have been observed to be divided upon specific issues, with different scambaiters engaging in activities that might be perceived more or less positively, activities which are encouraged by some and discouraged by others. Furthermore, from an ethical perspective, scambaiters have been observed to express differing philosophies when it comes to what is acceptable and what isn't as far as scambaiting activities go. In order to minimise the potential of negative impact upon specific YouTube channels or communities, the researcher will avoid linking specific practices to specific channels.

Although no specific comments or transcripts were identified as particularly sensitive, the subject matter nonetheless required careful handling. Instances of offensive or violent language, threats, and expressions of emotional distress were present within the transcripts. Accordingly, while the analysis did not exclude sensitive topics, it upheld scholarly integrity by avoiding gratuitous description and omitting unnecessary detail, employing content warnings where appropriate to mitigate potential distress. By balancing transparency with ethical responsibility, the study sought to navigate the complexities of data privacy and participant protection when working with publicly available material.

Another point of concern that is particularly pertinent given the subject of this research consists of considering the benefits that could be drawn from the findings of this work. It is evident to the researcher that online scammers are aware of and are actively engaging with the content created by scambaiters in attempts to avoid being impacted by such activities and to preserve their scamming operations. While there is no expectation that this piece of research will become popular reading among those who are actively involved in the committing of online fraud, the risk of this piece of research potentially equipping online scammers to better navigate the difficulties imposed upon their operations by scambaiters is something that has been considered in depth.

Throughout the data explorations conducted upon the video transcripts and the comment sections, several scambaiting strategies and techniques have been identified which need to be discussed as part of this piece of research. However, despite this data also being available publicly, the researcher will avoid going into too much detail about specific ways in which scambaiters manage to disrupt scamming operations, in order to minimize any educational value that could be gained by online fraudsters and applied to the improvement of their activities to become more resistant against modern scambaiting. Despite this, it is important to

note that, due to the data being used in this piece of research being publicly available, the likelihood of bad actors being aware of the most recent scambaiting strategies and techniques is not 0.

3.7 Challenges and limitations

While the methodology applied in this study provides a robust framework for analysing the phenomenon of scambaiting through the analysis of YouTube videos and comments, there are several inherent challenges and limitations that need to be acknowledged. These factors include the reliance on LLMs for data analysis, the constraints created by focusing solely on data from YouTube videos, and the issues around data quality due to the manner of retrieving it. Understanding these limitations is central to contextualizing the findings and their implications for the broader ecosystem within which scambaiting is enacted.

The employment of LLMs into the data analysis process has offered significant benefits in handling the qualitative data retrieved, particularly in terms of making it feasible to appropriately handle the extensive volume of text that has been captured. However, this reliance on AI-based tools introduces a number of challenges and limitations to be considered.

The research team recognized that while a human analyst is superior in interpreting nuanced, non-verbal, and contextual cues, the sheer scale of the dataset rendered a purely manual analysis unfeasible. The LLM was therefore employed as a tool for broad thematic identification across the entire dataset. The 5% manual sample of 215 videos served as the crucial, human-led validation of these identified patterns; the observations made during manual review were found to be predominantly consistent with the LLM's outputs, providing an additional layer of confidence in the reliability of the analysis.

While the capacity of the most recent LLMs to process and analyse large datasets effectively is significant, they are not immune to error. Despite their advanced capabilities, the outputs produced by LLMs can lack in contextual understanding and/or coherence, which can lead to inaccuracies in interpreting nuanced language, cultural references, and even specific terminology that is present in the scambaiting discourse. In light of this, some of the insights that were arrived at might have been generated without accounting for some of the elements just listed. Furthermore, the information available in the YouTube videos goes beyond the transcripts, with visual and audio queues often being utilised by scambaiters which would not have been captured in the transcripts processed. As such, the analysis conducted on the

transcripts is missing details that might have been present in the YouTube content but that could not logistically be included in the transcripts extracted.

Another concern to be considered is that LLMs have a capacity for hallucinating, generating outputs that contain factual inaccuracies arising from the model's constructions rather than actual data. While steps have been taken to combat the risk of hallucinations through prompt engineering and through random quality checks, risks still exist that could impact the validity of the analysis of results, potentially skewing the understanding of participants' comments and interactions. Such risks are however not believed to be significant considering the countermeasures employed.

LLMs are also prone to inherit and perpetuate biases present in their training data, which can lead to biased insights in the context of the scambaiting discourse. This is primarily a methodological limitation affecting the reliability of the findings. However, it extends into an ethical concern due to the researcher's responsibility to mitigate bias and ensure that the research does not inadvertently perpetuate harmful stereotypes present in the training data. At the time this research was conducted, the use of LLMs in academic research was a novel and contested practice, with scholarly opinions divided on their appropriate application and ethical implications (Hosseini et al., 2023; Perkins et al., 2023).

While LLMs are capable enough to facilitate the analysis of extensive data, the interpretation of outputs remains fundamentally reliant on the researcher's critical engagement with these results. By nature, qualitative analysis is subjective, which means that the interpretations can vary based on the researcher's knowledge, biases, and experience. The merging of LLM-generated insights with researcher-driven interpretations may introduce inconsistencies or skew findings if not thoroughly vetted.

Due to the use of YouTube-generated transcripts, inaccuracies and errors, especially with regards to language nuances, dialects, or technical language are likely to be present within the data. Such inaccuracies can affect the authenticity of the data and, consequently, the reliability of the analysis. While the capabilities of the LLM technology used in data analysis can navigate issues within qualitative data, the reliability of the analysis is still likely to be impacted when specific elements or contexts are not accurately picked up by the transcript generation.

Moving away from the topic of LLMs, another element to consider is the decision to limit the data collection strictly to content made available via YouTube. This decision has been taken primarily due to YouTube's significant popularity, causing it to be relatively reliable in terms of identifying and accessing video-form scambaiting content. However, throughout the research process the researcher became increasingly aware of how the activities of various scambaiters

and the communities that formed around them span across multiple platforms, each of them also being active on other platforms such as Reddit, Twitch, Discord, Patreon, TikTok, Ko-Fi, and others. As such, while all of the scambaiting operations identified are actively publishing content on YouTube and engaging with their communities via the features available, they are also leveraging other platforms. The researcher has explored some of the activity taking place on some of the other platforms mentioned above but due to significant additional data that would have been required to be collected in order to appropriately include these into the current research project, a decision was made to continue with looking strictly at YouTube activities. By centring the analysis on a single data source, there is a risk of the findings being skewed based on the inherent characteristics and practices of the platform, making the lack of cross-platform comparison a significant limitation to the study's generalizability and understanding of scambaiting as a broader phenomenon.

Another limitation to consider pertains to language and cultural barriers. The scambaiting content selected and analysed for this research predominantly occurs in English, with scammers often targeting potential victims in English-speaking countries, particularly the United States and the United Kingdom. However, online scamming operations extend beyond English-speaking regions, and it is likely that scambaiting activities are also occurring within non-English contexts. For instance, the content examined in this study occasionally discussed how scamming operations may target different geographical areas at varying times to align with the time zones of their intended victims. Consequently, a 'day shift' may focus on targeting local individuals, while a 'night shift' may concentrate on overseas targets. This research is therefore constrained to activities conducted in English.

The reliance on technology for the data collection and analysis processes are also to be considered, especially in relation to the reproducibility of this piece of research. Potential changes and variations in the technologies used (such as the YouTube API, the Python packages, and the version of OpenAI's LLM) can have a significant impact upon the ability to retrieve and process data consistently across future studies. The technological constraints also render the analysis susceptible to potential errors, bugs, or API limitations, highlighting the need for ensuring appropriate use of relevant technologies to minimize such risks. The researcher is confident that the technologies used for this piece of research have been utilised appropriately and every effort was made to mitigate potential errors, issues, and to adhere to API limitations.

While the methodology employed in this study yields valuable insights through the analysis of video transcripts and comments, the inherent challenges and limitations discussed above must be acknowledged. The reliance on LLMs for data analysis could prove problematic where the manual checks introduced are not sufficient to appropriately ensure quality, while the exclusive

focus on YouTube content limits the research's ability to capture the broader scambaiting ecosystem across various other commonly used platforms. These limitations reveal the importance of ongoing reflexivity within the research process, encouraging continuous critical engagement with the data and theoretical frameworks guiding the analysis. The recognition and addressing of these issues is hoped to enhance the credibility of this research project while contributing to a nuanced understanding of scambaiting practices in the digital world.

3.8 Chapter Summary

This chapter has provided an in-depth examination of the methodology employed to explore the phenomenon of scambaiting by analysing YouTube video transcripts and comments. The methodological approach integrates the theoretical framework of symbolic interactionism, bringing its own contributions providing unique perspectives and insights into the complexities of online scamming and the counteractions taken by scambaiters.

The significance of meaning-making processes and use of symbols and language is emphasized by symbolic interactionism bringing into focus the construction of identities carried out by the individuals navigating the scambaiting and online fraud landscapes.

The data collection and analysis is facilitated by various pieces of technology which enabled an efficient and standardized approach to handling the significant volumes of qualitative data retrieved. However, the technological elements are not without limitations. The reliance on LLMs for data analysis, while advantageous in processing vast amounts of data in a standardized fashion, raises concerns around accuracy and potential inherent biases in the outputs generated. The exclusive focus on YouTube content is also restrictive, failing to fully capture the broader dynamics of scambaiting activities that occur on other similar platforms. Additionally, the quality of the data has also included issues such as transcript accuracy and contextual nuances, which necessitates careful consideration when interpreting results.

Despite the challenges encountered, the chosen methodology offers a comprehensive framework for appropriately investigating the phenomenon of scambaiting in its complexity, and its implications within the wider arena of online fraud. Through critically engaging with the data and reflecting on the interplay of theoretical perspectives, the researcher hopes to contribute valuable insights to the understanding of this growing phenomenon. The findings will not only enhance academic discourse on scambaiting but will also inform future research towards continuing to expand the understanding of this rapidly evolving phenomenon.

The subsequent chapters will focus on presenting the analysis of the data collected and analysed, drawing upon the methodological foundations established in this chapter, as well as upon the literature previously discussed.

Chapter 4 The Performance of Scambaiting

4.1 Introduction

Scambaiting has emerged as a compelling and dynamic response to the pervasive issue of online fraud, characterized by its unique blend of humour, community engagement, and strategic resistance against scammers. At its core, scambaiting involves individuals known as scambaiters who purposefully engage with scammers, often impersonating vulnerable targets, to expose fraudulent practices and waste the scammers' time. This practice serves the dual purpose of disrupting criminal operations and raising public awareness about the tactics employed by fraudsters.

The rise of digital communication has transformed not only the ways in which people interact but also the nature of fraud perpetrated online. Consequently, scambaiting has garnered significant attention in these digital spaces, especially on platforms like YouTube and Twitch. Here, scambaiters document their interactions and share them with a broader audience, creating content that is simultaneously educational and entertaining. The evolution of digital technologies has equipped the scambaiting community with more sophisticated tools and channels of engagement, facilitating the expansion of their operations in a manner that mirrors

4.2 Strategies and Tactics Employed in Scamming and Scambaiting

In the realm of online fraud, the elaborate and often duplicitous nature of scamming strategies presents unique challenges for both victims and defenders. As scammers continuously refine their techniques to exploit vulnerabilities, so too do those determined to disrupt these malicious efforts adapt their own methods. This section explores the intricate web of tactics employed by scammers, ranging from emotional manipulation and social engineering to sophisticated technological tools, highlighting the psychological and operational dynamics at play. By dissecting the various strategies that characterize the world of scamming, we gain valuable insights into the underpinnings of this pervasive issue, setting the stage for a broader discussion on the countermeasures employed by scambaiters. Understanding these scamming strategies is essential not only for recognizing the signs of fraudulent activity but also for appreciating the ongoing battle between scammers and those who seek to protect potential victims.

4.2.1 Scamming Strategies

The intricate landscape of online fraud is populated by a myriad of scamming strategies that exploit human and technological vulnerabilities, which are used by individuals and organizations seeking to defraud unsuspecting victims. Scammers employ a plethora of tactics to manipulate victims, often leveraging psychological principles and social dynamics to establish a position of legitimacy and authority and to persuade their targets to comply with directions and instructions. One of the benefits that scambaiting brings to the fight against online fraud is the exposure and documentation of such tactics, which will be briefly explored in this section.

Most commonly, scammers have been observed to engage in tactics that rely on emotional manipulation, capitalising on fear, urgency, or sympathy to spur immediate action from victims. For example, phrases like “your account will be frozen” are frequently used to indicate that the target is at risk of suffering negative consequences unless they comply with the directions and instructions of the scammer. In one video, a scammer warns the target, "You can lose all the money in the bank, come back" (Kitboga, 2017a, 11:15), creating a sense of immediate financial peril.

The strategic creation of urgency aims to distance the target’s rationality, prompting them to act hastily and not giving them the chance to assess and evaluate their circumstances calmly. High-pressure tactics that create an atmosphere of panic are often leveraged in an attempt to deter the victim from scrutinizing the legitimacy of the requests being made. By inducing a heightened negative emotional state in their victims, scammers aim to reduce critical thinking and to promote compliance, thus rendering individuals more vulnerable to exploitation.

The narratives crafted and maintained by scammers frequently prey on the anxieties and concerns of their targets, often escalating emotional appeals until a tipping point is reached. Technical terminology and loaded language are frequently used to give credence to the narratives while playing to common concerns that victims might have. For example, hacking attempts made by foreign actors residing in China or Russia are frequently mentioned by scammers, playing on possible concerns around cybersecurity and the threat of foreign actors that the target might have. This was observed across multiple channels, with scammers making claims such as that a target's account had been accessed from Russia (ScammerRevolts, 2018, 13:18), that 'we found the hackers from the two states... Russia and another one is from China...' (DEYOCLUB, 2021b, 47:34), or that a fraudulent charge originated 'from Pornhub.com, from China' (More Kitboga, 2021a, 02:14:05).

Such elaborately constructed scenarios may include threats of legal action or claims of imminent financial loss, seeking to induce fear and panic. The emotional escalation, coupled with urgency, is also meant to persuade the victim to act “immediately” to avert dire consequences, fostering an environment where victims feel they have no choice but to comply. Such tactics reduce the likelihood of victims having the chance to conduct due diligence, while also strengthening the scammer’s position of perceived authority.

Across many scamming strategies, social engineering remains a prevalent tactic, involving the cultivation of trust and the establishing of legitimacy in order to ensure compliance from the victim. Scammers often disguise themselves as official representatives of reputable entities such as banks, government agencies, or tech companies, capitalising on societal norms that value authority. For example, in one Trilogy Media video, scammers impersonated agents from the 'US Government Grant Department' (Trilogy Media, 2021). Thus, strategic impersonation, where successful, aims to reduce the victims’ scepticism and to ensure compliance. Furthermore, through data breaches, social engineering, or poor cybersecurity practices, scammers are also able to get sensitive information which they can leverage to further establish legitimacy and credibility.

4.2.2 Scambaiting strategies

In direct opposition to the deceptive tactics employed by online fraudsters, the practice of scambaiting has emerged as a form of digital vigilantism. Scambaiters proactively engage with scammers, not as unsuspecting targets, but as adversaries with the intent to disrupt their fraudulent activities. Intriguingly, many of the strategies employed by scambaiters are a direct mirror of the scammers' own methods, representing a strategic reversal of deception and manipulation. This symmetrical approach of "fighting fire with fire" involves leveraging the very tools of social engineering, technological exploitation, and psychological pressure that define the scammer's playbook. The primary objectives of these strategies are to waste the time and resources of scammers, collect information that can be passed on to law enforcement, dismantle their operations, and educate the public about the persistent threat of online fraud.

One of the most foundational strategies in scambaiting is the intentional consumption of a scammer's time and resources, a tactic built upon a foundation of elaborate deception. Just as scammers construct false identities as representatives of legitimate entities, scambaiters create equally fictitious personas, often characterized by extreme gullibility or eccentricity. They feign technological ineptitude, fabricate convoluted backstories, and engage in prolonged, seemingly pointless conversations to keep the scammer occupied under the false pretense of

an imminent payout. This mirrors the scammer's own use of crafted narratives to establish legitimacy, but inverts the goal: where the scammer deceives for financial gain, the scambaiter deceives for resource attrition, thereby increasing the operational cost and reducing the profitability of the fraudulent enterprise.

This parallel use of social engineering is even more pronounced in the scambaiter's efforts to gather intelligence. In a direct reflection of how scammers cultivate trust to extract sensitive information from victims, scambaiters build a false rapport to coax scammers into revealing personal details, operational locations, and financial accounts. This investigative approach seeks to collect actionable information that can be reported to the authorities. The psychological manipulation at play is reciprocal; both parties are engaged in a performance, attempting to convince the other of their authenticity in order to achieve their clandestine objectives. The key distinction lies in the ethical framing and ultimate goal: the scammer manipulates for criminal exploitation, while the scambaiter manipulates for disruption and justice.

Furthermore, the leveraging of technology presents another area of strategic symmetry. Scammers frequently use tools like Remote Access Trojans (RATs) to gain unauthorized control over a victim's computer, enabling them to steal banking information or install ransomware. In a strategic reversal, the more technically proficient scambaiters deploy the very same class of tools against the fraudsters. By tricking a scammer into installing a RAT on their own machine, a scambaiter can gain access to the scammer's system to gather evidence, delete victim lists, and actively sabotage their operations from within. This appropriation of the scammer’s technological arsenal represents a sophisticated form of counter-attack, turning the weapons of digital fraud against their perpetrators.

Finally, a significant aspect of the scambaiting movement is its focus on public education and awareness, which serves as a broader counter-narrative to the scammers' deceptions. Many prominent scambaiters utilise platforms like YouTube and Twitch to broadcast their interactions, deconstructing scams in real-time. By exposing the specific scripts, emotional manipulation tactics, and technological tricks used by fraudsters, they empower the public with the knowledge to recognize and avoid these schemes. This educational outreach effectively inoculates potential victims against the very strategies of urgency and fear detailed in the previous section, creating a more resilient and informed online community better equipped to defend against the ever-evolving landscape of online fraud.

4.3 Technological Tools and Manipulations

In an ever-evolving landscape, the battle between scammers and scambaiters demonstrates a dynamic interplay of technology and strategy on both sides. While scammers employ a range of digital tools to perpetrate their fraudulent activities, scambaiters are equally resourceful in utilising advanced technologies to counter deceitful practices and disrupt operations. This section delves into the various technologies and strategies employed by both scammers and scambaiters. By examining how scammers leverage technology to achieve their goals and how, in response, scambaiters develop and employ countermeasures, we can gain a deeper understanding of the ongoing cat-and-mouse game that unfolds in the digital realm, underscoring the significant role technology plays in this modern-day warfare against online fraud.

4.3.1 Technologies and strategies used by scammers

Technological tools and platforms also feature heavily in the arsenal of online scammers. Some of the most common technologies used by scammers involve remote access tools (such as TeamViewer, AnyDesk, Supremo, etc.), Voice over IP (VoIP) software, and call spoofing technologies.

From a dramaturgical perspective, these technologies function as essential 'props' and elements of the 'setting' that bolster the scammer's front-stage performance. In this context, scammers, as actors, utilise these technological props to construct a credible performance for their audience, the victim. Call spoofing and VoIP, for instance, manipulate the digital setting by creating the illusion that the scammer is calling from a legitimate and trustworthy location, such as a local number or the official line of a well-known corporation. Remote access tools are particularly powerful props; by guiding a victim to install and run legitimate software like TeamViewer, the scammer enacts the role of a professional technician, thereby borrowing the software's own legitimacy to enhance their performance. The successful deployment of these props is crucial for managing the victim's impression, reinforcing the scammer's assumed role, and making the fraudulent scenario appear authentic.

Furthermore, upon gaining remote access to a victim’s computer, scammers also leverage existing functionality from the victim’s systems to further establish legitimacy or to cause damage. Scammers have been frequently observed to use syskey (a security feature available on Windows that has been discontinued in more recent versions) to lock a victim’s computer and take it hostage, as demonstrated in a video by Malcolm Merlyn where a scammer uses the

tool in retaliation (Malcolm Merlyn, 2018). Similarly, scammers have also been observed to use the Command Prompt to fake various computer issues in order to fabricate an issue that they can support with.

More advanced scamming strategies involve the use of more customised tools, platforms, and websites. These are analysed here as meticulously constructed Goffmanian "settings" crucial for performance credibility. For example, scammers create functional websites that they use in various ways to defraud victims, whether by creating fake online shops and selling products without delivering them, or by creating a front for the fraudulent operations that they run. Some investment scams have been found to have custom investment platforms, which present a highly credible setting by showing real-time market activity while faking the returns of the investments defrauded from their victims, so as to maintain a front of legitimacy and profitability.

Additionally, scammers have also been observed to leverage existing platforms and free tools to enhance their operations. This can be in the form of fake reviews for the fake services they provide via reputable platforms (such as Trustpilot), the use of Google forms to simulate legitimacy by getting their target to provide requested pieces of information, or the use of online demo bank accounts under the pretence that they are real, just to list a few examples.

From an operational perspective, scamming operations often appear to function in a hierarchical and organised fashion, akin to a business, where different scammers play different roles and have different responsibilities within the perpetrating of online fraud. This was evident in a lengthy Kitboga video where the initial scammer repeatedly had to consult with and eventually transfer the call to his 'manager' (More Kitboga, 2021a). This hierarchical structure is a common tactic, with scammers in other interactions also escalating calls to a supposed 'senior technician' to project authority (SkeletonSyskey, 2024). This is beneficial towards adding an element of authenticity with scammers transferring their “customers” to supervisors and managers reflecting processes similar to real businesses.

Additionally, as some scambaiters have noted, this is also a security measure for the scamming operation, where the segmentation of information and the protection of sensitive information (such as bank accounts, or victim data) is central towards ensuring individual scammers are not able to steal this information and engage in scamming outside of the “organization”.

Finally, while the primary objective of scammers is generally the extraction of funds from their targets, in instances where significant access to the victim is established scammers are able to leverage that as a resource to facilitate subsequent scams. For example, scammers might make use of victims’ bank accounts to launder money, or get them to perform various tasks which

contribute to fraudulent schemes, such as receiving and forwarding packages which could potentially contain funds extracted from other victims, or illicit goods. Furthermore, by using victim details, scammers might attempt to create various accounts in their name, which could be used to launder money. This is relevant for example to cryptocurrency exchanges where KYC (know-your-customer) practices often require some form of identification, which is one of the items scammers might try to get from their victim. In addition to further enabling scamming operations, involving victims in fraudulent schemes also adds a layer of protection for scammers.

As it can be seen from the brief exposition above, the variety of strategies and tools available to scammers is vast and varied depending on the scamming operation looked at. However, through dissecting and documenting scamming operations, scambaiting is able to capture not only the methods used by scammers, but also how these evolve over time as technology advances and new scams are being developed.

4.3.2 Scambaiting Strategies

Scambaiters have a wide array of tools and strategies available that they employ to combat scamming operations. Before expanding on these, it must be noted that, based on the video content observed, two primary categories of scambaiting can be distinguished: investigative scambaiting and engagement-focused scambaiting. This distinction was apparent across the sample of channels analysed. For example, the content of channels like Jim Browning's frequently exemplifies the investigative approach, focusing on the technical infiltration of scammer operations and the gathering of evidence for authorities. In contrast, channels such as Kitboga's epitomize the engagement-focused model, with videos consisting of long-form interactions centered on elaborate role-playing and wasting scammers' time through humorous scenarios. While a precise quantitative breakdown is beyond the scope of this analysis, both models were well-represented in the data, with many channels blending elements of each.

Investigative scambaiting content is frequently shorter in form, consisting primarily of scambaiters presenting the findings of investigations conducted largely offline. The focus for investigative scambaiting falls primarily on dismantling and exposing the inner workings of scam operations, including sensitive details such as names and pictures of scammers, as well as their geographical locations and company names. This investigative angle is characterized by a somewhat less transparent methodology, with scambaiters withholding certain techniques for ethical considerations and to maintain a strategic advantage over scamming operations. A significant benefit of this approach, however, is its capacity to deliver actionable intelligence,

which is frequently passed to the appropriate law enforcement authorities, and to gain a higher level of access than could be achieved through direct interaction alone.

In contrast to the investigative model, engagement-focused scambaiting centers on the live interaction between the scammer and the scambaiter. Such content is usually longer in form and centres on the dynamic exchanges that occur between the two parties. It is structured around humorous and engaging interactions that highlight the strategies scammers use to lure victims and how scambaiters counteract them. This type of content thrives on the spectacle of engagement, often resulting in entertaining narratives that capture the absurdity of scammers' tactics and enable the audience to witness real-time interactions with nefarious actors. While a significant level of opaqueness is maintained in its operational aspects, this approach is often more transparent than investigative scambaiting in showcasing the practical techniques and strategies employed by scambaiters to manipulate and frustrate scammers. Although emphasis remains on the disruption of scamming activities, effort is also put towards creating a captivating spectacle for the audience, often incorporating elements of comedy and creativity to entertain and educate viewers.

It is essential to note that, despite the distinction made above, scambaiting content does not generally fit neatly within one category. Rather, it exists on a spectrum between the two, combining elements from both in different measures. Investigative scambaiting often includes direct interactions with scammers, and similarly, engagement-focused scambaiting benefits from rigorous investigation. Different channels, and even different pieces of content, will rely more heavily on one approach or the other, utilising whatever blend is most suitable for the targeted scamming operation and the scambaiter's specific goals and style.

This distinction between "investigative" and "engagement-focused" scambaiting is central to the main argument of this research. "Engagement-focused scambaiting" is argued to be the clearest manifestation of entrepreneurial digilantism, as its methods are directly driven by the need to create entertaining, monetisable content. The performance, the humour, and the live interaction are not merely tactics for disrupting scams; they are the core product that sustains the scambaiter's operation through audience engagement and monetisation.

Beginning with a focus on strategies more prevalent in engagement-focused scambaiting, one of the most commonly encountered tools is the use of humour and absurdity. Humour is frequently observed as a powerful tool in disarming scammers, employed by scambaiters through the introduction of absurd scenarios or exaggerated character portrayals. The entertaining dynamic that emerges within these interactions challenges the scammer's authority, disrupting their intended narrative while simultaneously creating entertainment for the audience.

Role-playing is also heavily used by scambaiters to engage scammers in order to gather intelligence, expose strategies, or simply waste their time. Most frequently, scambaiters will take up the role of a victim to lure the scammer into a false sense of security, making it possible to more effectively disrupt their operations. From a dramaturgical perspective, this adoption of a victim persona is a carefully constructed "performance." The scambaiter embodies a "role" defined by characteristics that scammers are conditioned to seek: vulnerability, technological illiteracy, loneliness, and gullibility. This performance is a strategic exercise in impression management, where the scambaiter, as the actor, meticulously controls the "front" they present to their audience, the scammer. By feigning confusion, displaying emotional vulnerability, and adhering to the expected script of a victim, the scambaiter manipulates the scammer into believing they are in control of the interaction. This successful performance allows the scambaiter to define the situation on their own terms, creating an environment where the scammer's own tools of manipulation can be turned against them.

However, on occasion, scambaiters might also role-play as a third-party character, such as a banking official or a customer support representative, to provide legitimacy to the scenario being portrayed or to frustrate the scammers' attempts at meeting their goals. For instance, scambaiters were observed adopting third-party personas, such as Kitboga pretending to be a customer support agent asking a scammer, "It looks like your longest hold time was 1 hour and 10 minutes, are you willing to beat that record today?" (Kitboga, 2021a, 15:05). In other instances, scambaiters might pretend to be opposing scammers to create competition for the scammers being targeted, trick them into a particular narrative direction, or disrupt their plans. Quite amusingly, it has been observed that in instances where scambaiters introduce competing scammers that are closer to defrauding the target than the actual scammers, this creates a similar type of urgency and emotional engagement as what scammers generally try to instil in their targets.

To effectively impersonate these various roles, scambaiters often use a variety of technological solutions, including voice-changing hardware/software, VoIP systems, and audio recordings to provide background or environmental noises where needed. This was frequently observed in the data, with creators like Kitboga using voice modulation software to switch between multiple characters in real-time to confuse a scammer, while Jim Browning has used similar technology to role-play as a character to navigate a scammer's script (Kitboga, 2020; Jim Browning, 2021), and others using it to enhance their victim personas (Rinoa Antidote, 2021). In Goffman's framework, these technologies function as essential "props" and elements of the dramaturgical "setting," which are crucial for the credibility of the scambaiter's front-stage performance. Voice modulators, for example, are part of the "personal front," allowing the actor to align their vocal characteristics with the chosen role, such as an elderly woman or a stern bank manager.

Soundboards that play background noises, a barking dog, a television, or street sounds, help construct a believable setting, adding a layer of authenticity that makes the performance more difficult to discredit. These props are not merely accessories; they are instrumental in managing the scammer's impression of the situation. By creating a coherent and convincing scenic backdrop, these tools prevent interactional mishaps that could expose the performance as a fabrication, thereby ensuring the scambaiter maintains control over the definition of the reality being presented.

Another frequently observed strategy consists of role reversal to frustrate scammers' schemes. Initially, scambaiters generally begin the interaction by posing as a prospective target, seeking to convince the scammer of their legitimacy and viability as a victim. However, once the scambaiter has convinced the scammer of their legitimacy, they can begin to transition into a more confrontational or non-cooperative role, challenging the scammer's authority and reversing the power dynamics. This tactic enables scambaiters to assert control, using the scammers' assumptions against them while illustrating the absurdity of the scam.

Technologies and strategies used by scambaiters

Technology is an essential element in the scambaiters' toolboxes, with a significant variety of tools, systems, and platforms being leveraged towards combatting online fraud. Some of these relate to preserving the personal security of scambaiters and their systems. Engaging with online scammers involves a significant amount of risk for scambaiters if they are not careful about protecting their identities and their systems, especially in instances where they are also livestreaming their activities. To this end, the use of VPNs and virtual machines can be framed as the essential architecture of the scambaiter's "backstage," allowing safe performance preparation hidden from the front stage. Additionally, the use of VoIP systems allows scambaiters to not only more effectively enable their role-playing but also to have an additional layer of protection by not disclosing personal telephone numbers.

Tools such as remote access software, as well as others that are not openly disclosed by scambaiters, are frequently used in a more offensive fashion, enabling scambaiters to gain access to scammers' computers and systems. While challenging, this often proves quite fruitful since being able to access scammers' systems enables scambaiters not only to retrieve valuable intelligence about the scamming operations but also to cause damage by deleting files containing victims' details, and even set up surveillance. A notable example of advanced disruption is the use of 'call flooding', where scambaiters employ specialized software to

overwhelm a scam call center's phone lines with automated calls, effectively shutting down their operations as demonstrated by Malcolm Merlyn (2021, 03:00).

On the latter point, scambaiters have on occasion been able to use the remote access they had to scammers' systems to identify and disrupt active scamming attempts by alerting either the victim (where their contact details were available) or relevant third parties/authorities (such as banks or local police departments). The investigative work enabled by such levels of access has also been observed to be quite valuable since scambaiters have been able not only to reveal the identities of those engaging in online fraud but also to collect compelling evidence of wrong-doing and submit it to appropriate law enforcement bodies.

Scambaiters have also been observed to design and implement their own pieces of technology towards enhancing their ability to convince scammers of the authenticity of the persona played, or simply to frustrate scammers' plans and attempts to complete certain objectives. Such examples include fictitious banking websites that appear legitimate enough for scammers to believe they are real, as used by Rinoa Poison (2021), and custom-built honeypot websites designed to trap scammers in an 'Impossible Maze' (Kitboga, 2021a). Such tools serve to waste scammers' time and gather intelligence on their methods.

From a dramaturgical perspective, these technological tools function as essential 'props' and components of the 'setting' that are indispensable for the scambaiter's front-stage performance. Technologies like VoIP systems and fictitious banking websites are carefully selected props that lend an air of authenticity to the scambaiter's portrayal of a gullible victim or any other chosen persona. The virtual machine acts as the stage itself, a controlled setting where the drama unfolds, insulated from the scambaiter's real-world backstage reality. The successful management of these props is crucial for 'impression management', as a single technical flaw or an unconvincing prop could shatter the illusion, leading to the failure of the performance and potentially exposing the scambaiter to risk.

Finally, scammers themselves are on occasion a valuable resource for scambaiters. Some of the content that has been observed includes instances where scammers have chosen to collaborate with scambaiters and to provide sensitive information about the scamming operation they are part of, as well as about the individuals involved, thus assisting scambaiters with their efforts to investigate and disrupt scamming operations. A well-known example is a video by Jim Browning featuring an insider from a 'pig butchering' scam who provided crucial information, including CCTV access, to expose the entire operation (Jim Browning, 2022). Furthermore, some scammers have even gone as far as recording and providing audio/video footage of the inside of scamming call centres towards aiding scambaiters in their efforts. This

is an interesting phenomenon that deserves additional attention, but it will not be touched upon in detail in this piece of research.

4.4 The Interplay between Scammers and Scambaiters

The interactions between scammers and scambaiters present an interesting display of dynamic exchanges characterized by continued shifts and role reversals. This section delves into these interactions, hoping to shed some light on how the power dynamics, identity construction, and strategic engagement unfold during scambaiting encounters.

Predominantly, scambaiting interactions commence with scambaiters responding to the initiation of a scammer, most frequently taking the form of a phone call. This call is either made by the scammer to the scambaiter to initiate the scam, or by the scambaiter to the scammer as a response to an initiation (e.g. pop-up ads, voicemails, social media ads, etc.). This was the most common pattern of initiation observed across the vast majority of videos analysed. For instance, in numerous videos from channels like Scammer Payback and Kitboga, the interaction begins with the scambaiter dialling a number found on a fake pop-up ad or in a phishing email, thereby responding directly to an initiation attempt.

The scammer presents themselves as a representative of a legitimate organization, deploying tactics of emotional manipulation, urgency, and scripted dialogue exchanges, designed to exert authority and to ensure compliance. This places the scammer in a position of power, often guided by scripts and by their knowledge of the scam they are perpetrating. For example, scammers often create scenarios where their target needs to urgently comply with their instruction or risk suffering serious consequences, with scammers preying on the fear and panic they are trying to instil in the potential victim. Other times, scammers might initiate scenarios where their target has something to gain from the interaction (e.g. a fictitious refund for a product/service) with scammers leveraging the target's desire to make this gain to lead them into a scenario where a justification can be made for the target to make a payment to the scammer. In most scenarios, the target is expected to adopt a passive role, adhering to the scammers directions and instructions leading to a linear narrative within which the scammer is trying to keep the victim progressing forward.

This power dynamic begins to shift when scambaiters begin challenging the authority of the scammer. Scambaiters employ various strategies such as adopting the role of a confused or distressed victim, putting the onus on the confusion and distress for their inability to strictly follow the scammers instructions, and carefully navigating the emerging dialogue to keep

pushing against the exerted authority. This is a dexterous endeavour as scambaiters need to find the appropriate balance so that scammers' activities are sufficiently frustrated without causing them to disengage. The narrative deviations that scambaiters trigger lead to the fracturing of the scam script, revealing the vulnerabilities of the scammer's constructed identity, as well as the flaws within the scenario created.

From a dramaturgical perspective, the scammer's performance is entirely dependent on a pre-established 'script.' The scambaiter, however, refuses to be a passive audience member. They intentionally deviate from the script with a deliberate attempt to force the scammer "out of frame" or to "break character." By introducing unexpected variables, such as "feigning technological incompetence," asking tangential questions, or injecting absurdity, the scambaiter introduces elements that the scammer's rigid framework cannot accommodate. This subversion forces the scammer off-script and into the realm of improvisation, a skill for which they are often ill-equipped. The fragility of the scammer's performance is thereby exposed; their authority is not inherent but a carefully constructed artifice that shatters the moment their scene partner ceases to play along. The resulting breakdown of the performance reveals the inauthentic self behind the role.

This is frequently accompanied by significant frustrations expressed by scammers who are forced to either attempt to restore the narrative of their script, or to improvise and attempt to recuperate the authority to control the direction of the interaction. Scambaiters also use various prompts to reveal the deception in the scammers narrative and identity by concealing them in conversation. A common example is found in scambaiters inquiring about the scammers location by using leading questions which do not make sense geographically, but with which the scammers, not having the geographical knowledge that their character should have, play along. A good example of this would be a scammer who claims to live in California, being asked about if he gets to visit the Golden Gate Bridge frequently since he lives so close to it. The way in which the question is posed feeds misinformation to the scammer who, by failing to pick up on it, ends up betraying their ignorance as to the area within which they claim to be residing.

Simultaneously to scambaiters attempting to find the appropriate level of engagement with the scammers so that they can fulfil their objectives, scammers are also probing their target to determine their legitimacy, and the potential they have towards being defrauded. This is evident when scammers express their doubts directly, with one noting, 'I think this guy is very suspicious. He tried to log in without any credentials...' (More Kitboga, 2021b, 30:07), or when their tactics shift dramatically upon discovering their target has no money (Kitboga, 2022a). The tactlessness and clumsiness with which this question was posed and the scammer's evident intention of ensuring that their target is able to fulfil a task that is included in the script at a later

point in time create an amusing conflict between the scammer's attempts at being covert in their goals yet so obvious through their manner of engaging.

As the interaction progresses, the scambaiter's humour and assertiveness enable them to reclaim agency, leading to a role reversal where the scammer is no longer in control. This shift often occurs when the scambaiters shifts towards a more confrontational approach, using absurdity and mockery to challenge the scammer's tactics. The power reversal is pronounced as the scambaiter's confident engagement forces the scammer to navigate a context of uncertainty.

Scammers who typically rely on control and authority become flustered and defensive when faced with consistent resistance to the narrative they are trying to push, which not only disrupts the intended flow of the scam, but also situates the scambaiter as an active participant in the confrontation. Frequently, this prompts the scammer to resort to aggressive or bewildered responses. As power dynamics shift, scammers often exhibit visible signs of anger, frustration, and even bewilderment, having been observed to suffer from strong emotional outbursts, with one exclaiming, 'I'm gonna curse you now. I'm gonna f***ing curse you now' (Scammer Payback, 2021a, 02:22). Attempts to regain authority through threats and aggressive language become more likely to occur and their efforts to regain control are becoming increasingly desperate.

This contributes to the fracturing of their identity as a figure of authority and expertise (such as a support technician), further revealing the deception behind their claims and the desperate need to dominate the narrative despite increasing confusion. In response, scammers may resort to confusion-inducing tactics such as abruptly shifting the conversation back to vague technical requests or legal ramifications in a bid to regain the upper hand, which underscores their struggle to maintain the fictitious identity as the narrative blurs the line between authority and vulnerability.

The ongoing interactions are characterized by their fluidity, with control rapidly oscillating as each party adapts to the other's tactics. The dance between the scammer and scambaiter manifests as an intricate negotiation of power, where moments of authority can easily shift back and forth. Each individual's responsiveness to verbal cues, emotional expressions, and audience feedback plays a crucial role in this dynamic interaction. For example, scambaiters might note a scammer's defensiveness and adapt their approach accordingly – escalating humour to destabilize the scammer further or adopting a more serious stance to maintain the authenticity of the character played. This back-and-forth engagement occurs with roles continuously reshaped exemplifying the complex nature of social interactions where authority and identity are transient, being sculpted and adjusted to best fit the needs of whoever holds the character.

The authenticity of the interaction is also enhanced through various technological solutions employed by scambaiters. Pieces of hardware and software are frequently used to provide an additional layer of authenticity through simulating various other elements of what a real interaction would look like. For example, scambaiters could use audio recordings of various environmental factors to simulate specific scenarios and occurrences. Where a scammer instructs the scambaiter's character to drive to a store, the scambaiter could use audio recordings of a car engine starting up or audio of traffic to maintain authenticity.

Similarly, scambaiters also use technology to manipulate specific elements from the virtual environment within which interactions happen. Scammers frequently attempt to gain remote access to their target's computer seeking to retrieve information and also to reinforce their control over the situation. Scambaiters have been observed to utilise software to simulate functionalities that they are unable to legitimately employ for security reasons. For example, scammers might try to access the computer's video camera to verify the authenticity of their target. To combat this, scambaiters could set up the camera so that when it is accessed, it displays footage that could mislead the scammer into confirming the authenticity of the target. More recently, scambaiters have been leveraging generative AI technology to apply filters so that, when captured by the camera, their appearance is altered so that their identities remain concealed, while being able to react to the scammer in real time.

Technology also plays a significant role in the interactions from a goals perspective with scammers and scambaiters both seeking to leverage technological channels to meet their objectives. Scammers might try to gain access of a target's computer so that they obtain access to various accounts, or so that they can use it for ransom by changing passwords. Scambaiters on the other hand might seek to access scammers systems so that they can download/delete files, retrieve identifiable information, or even use the systems to further extend the access gained and to set up surveillance mechanisms.

For scammers, the objective of gaining remote access is much simpler to achieve since this is frequently a pre-requisite for them to go through their script. Because of this, scambaiters need to provide this access or risk the scammer disengaging due to not being able to proceed. However, even though access is gained by the scammer, what is being accessed is always a virtual machine carefully crafted to appear legitimate, making any damage that the scammer could inflict, inconsequential, since they are not interacting with a real computer.

On the other hand, scambaiters gaining remote access to scammers' systems is much more difficult since this is something that would trigger significant red flags for a scammer, so it needs to be conducted covertly. However, once successful, scambaiters achieve a strong position

from which they can cause significant disruption and retrieve valuable intel that can be used most importantly to identify scammers and to identify potential victims and warn them.

The dynamic interactions between scammers and scambaiters portray a complex interplay of role reversals, power shifts, emotional engagement, and contextual adaption, facilitated through role-playing, humour, and technology. Identity and authority are constructed, challenged, and transformed through these encounters with both the scammer and scambaiter engaging in negotiating behaviours to simultaneously pursue multiple goals. Scammers seek to build authenticity and legitimacy while at the same time probing their target to ensure their authenticity, with scambaiters in turn similarly building the authenticity of their own characters while simultaneously seeking to dismantle and disrupt the operations of the scammer. Scambaiter's unique positioning enables them to reclaim agency against deceitful practices, creating an engaging narrative that underscores the complexity and fluidity of roles. Ultimately, it is important to note that scammers never have any real control in their interactions with scambaiters with any apparent control being afforded and enabled by the scambaiters so that they can meet their goals and objectives.

4.5 Identity Construction

In delving into the intricate dynamics of identity formation within the realm of scambaiting, the fluidity and complexity embedded in the interactions between scammers and scambaiters can be observed. Central to these exchanges is the strategic construction of identities, where scammers meticulously curate authoritative personas aimed at engendering trust and compliance among their targets. Conversely, scambaiters exhibit remarkable adaptability, oscillating between various roles (such as naïve victim, knowledgeable protector, and educator) while simultaneously challenging the scammers' narratives.

This section will explore how these multifaceted identities not only reflect personal motivations and emotional reactions but also shape broader perceptions of victimhood and agency within the digital landscape. By illuminating the performances undertaken by both parties, we gain insights into the ethical considerations inherent in scambaiting, the impact of community engagement, and the ongoing dialogue around the nature of fraud and exploitation in contemporary society.

Identities Formed

The interactions between scammers and scambaiters illustrate the complex nature of identity construction within these engagements. The identities created and navigated are shaped by the conditions and contexts of the exchanges that take place and are imbued with symbolic meanings that reflect broader societal attitudes toward fraud, authority, and perceived victimhood. This section aims to dissect the multifaceted identities of both scammers and scambaiters, exploring the dynamics that unfold during the interactions observed.

Predominantly, scammers are observed to construct their identities around authority and legitimacy. By adopting roles as representatives of reputable organizations, such as tech companies or government agencies, they create a façade intended to establish trust and compliance among potential victims. To reinforce this authoritative aspect of their identity, scammers frequently employ technical jargon and scripted narratives. The use of specialised terms in technology and finance provides the illusion of expertise, assisting the scammer in establishing themselves as knowledgeable, thereby creating a false sense of security for the victims. The use of premeditated scripts also contributes by enabling scammers to include elements of emotional manipulation, generally urgency or fear, into the narrative, helping to further solidify the authoritative persona.

Interestingly, within these authoritative personas, scammers also tend to portray their characters as hard-working, family-oriented, and financially well-off. Data from interactions shows scammers making claims such as working 'two jobs' (Kitboga, 2021d, 24:17), being a 'software engineer' at Microsoft (Kitboga, 2022b, 27:35), having 'graduated from Oxford and... did masters from Harvard' (Kitboga, 2021c, 01:22:48), working long hours (Kitboga, 2021c, 39:14), and having 'pets and... children' to support (Kitboga, 2021c, 32:15). Through scambaiters' probing into the personas created by scammers, asking about their jobs, families, or other elements that could betray the fraudulent nature of the character, it has been often observed that scammers seek to establish their character in alignment with the expected norms of their victims. This frequently results in comedic circumstances where the attempts of the scammer to embody the norms they believe their character should have comes into conflict with their ignorance around what such norms entail. For example, when asked about their work schedule, scammers have, on occasion, indicated ridiculous shift lengths and patterns, coming into conflict with what the standard practices would be within their respective field. Claims of 6-day working weeks of 14-hour shifts, while attempting to portray the scammers as hard-working individuals, end up being detrimental to the authenticity they are trying to build around their personas.

Despite the authority projected, the scammer's performance is inherently fragile. When confronted by scambaiters, who often employ humour and absurdity as tactics, the authenticity

of these constructed identities begins to unravel. Many scammers experience cognitive dissonance regarding their actions, an internal conflict arising from their attempt at establishing a self-image as a competent professional becoming juxtaposed with the reality of their fraudulent behaviour. As scambaiters mock the scammers' narratives, this internal conflict becomes apparent, exposing the underlying anxiety and shame that contradict the successful image they strive to project.

This pressure to maintain an authoritative persona when challenged leads to emotional defensiveness, often manifesting as aggression or evasiveness. The inability to maintain a scripted identity reveals the tenuousness of their assumed authority. This confrontation not only disrupts the scam narrative but also signals the vulnerability behind the scammer's constructed identity, leading them to resort to threats or evasive strategies in an attempt to regain control of the interaction. The exposure of the undercurrent of anxiety, fear, and even guilt on rare occasions, triggers a shift in identity from authority to desperation, which, as it becomes more salient, pushes the scammer to emotional outbursts and to increasingly desperate attempts at recovering their previously established role. This dynamic is clearly demonstrated in interactions that bring the scammer to a point where repeated failures in stealing any money, drive them to resorting to insults and claims they knew it was a bait all along (Kitboga, 2022c). This is occurring even in the absence of any reveals, appearing to enable the scammer to save face by placing the burden of their failure not on their inability to succeed in the scam, but on there not having been any chance of success in the first place. In instance when such failures occur, in the absence of a reveal, scammers have been observed to revisit their targets again and again over a large period of time, making another attempt at defrauding them (Kitboga, 2022d)

Scambaiters, on the other hand, navigate a spectrum of identities throughout their engagements, frequently shifting between the roles of naïve victim and knowledgeable protector. Initially assuming the guise of a confused victim, scambaiters draw scammers into a false sense of security, allowing them to further the engagement and gather vital information about the scam.

From a dramaturgical perspective, this initial persona is a meticulously performed 'role' on the 'front stage' of the interaction. The scambaiter engages in sophisticated 'impression management' to project a convincing 'personal front' of a vulnerable individual. The key characteristics of this role include technological incompetence, excessive politeness, confusion, and a fundamental trust in the scammer's authority. The scambaiter's performance of this role serves to manipulate the interaction by confirming the scammer's expectation of an

easy target, thereby managing the scammer's impressions and lulling them into a state of overconfidence where they are more likely to make mistakes or reveal crucial information.

To support this dramaturgical analysis, specific examples from the data illustrate this type of 'performance' and 'impression management' in action. For instance, in one notable case, a scambaiter playing the part of a technologically illiterate senior citizen spent over an hour feigning an inability to locate the "any" key on their keyboard, a classic performance of incompetence designed to frustrate the scammer and waste their time. In another interaction, a scambaiter, instructed to purchase Google Play gift cards, instead repeatedly purchased greeting cards, meticulously describing each card's sentimental message to the increasingly agitated scammer while claiming to misunderstand the instructions. A third example of such impression management involves a scambaiter who, after being told their bank account was compromised, expressed effusive and tearful gratitude to the scammer for "being a hero" and "saving" them, thereby reinforcing the scammer's ego and keeping them committed to the prolonged engagement.

As they gain traction in the interaction, scambaiters pivot from this victim persona to adopt a challenging stance, using humour to interrogate the scammer’s identity and disrupt their authoritative façade. This ability to oscillate between roles not only empowers scambaiters but also helps reclaim agency over the interaction, marking a significant shift in power dynamics. Alongside the identities of victim and challenger, scambaiters also adopt the identity of educators, aiming to inform both the scammer and their audience about the realities of fraudulent behaviour, with some content being explicitly framed as a way to reveal information that 'Scammers Don't Want You To Know' (The Hoax Hotel, 2021). For the audience, the interactions often feature small pauses from the engagement, within which the scambaiter will draw attention to specific elements, providing explanations and insights into the mechanisms behind the scam. This strategic approach allows scambaiters to reinforce their moral superiority while exposing scammers’ tactics.

By narrating their experiences, scambaiters construct a collective identity centred around advocacy and resistance, transforming individual actions into a broader mission against exploitation. The transitions of scambaiters into the role of educator signal a deliberate choice to leverage their knowledge and experience for public good, reinforcing a narrative that positions them as advocates for consumers in the digital age.

As scambaiters and scammers engage, the transitions of identity experienced by both parties highlight the fleeting nature of authority and victimhood in these interactions. Scammers, while aiming to embody power, often find themselves grappling with unmasking emotional honesty as they struggle to maintain their deceptive narratives. Simultaneously, scambaiters craft and

adapt their identities to not only fulfil their roles in confronting fraud but also to resonate with the audience, fostering a sense of solidarity and shared purpose among viewers. Through this complex interplay of identities, scambaiters not only challenge scammers but also affirm their own roles as protectors and advocates within the digital ecosystem.

4.5.1 Construction of Identities

The interactions between scammers and scambaiters reveal intricate constructions of identity and the negotiation of roles within these engagements. At the forefront of the scammers' tactics lies an effort to project authority and legitimacy. Scammers often adopt identities associated with reputable organizations, presenting themselves as technical support representatives or agents of law enforcement. This masked persona is strategically designed to instil confidence in their targets, utilising scripted dialogues filled with technical jargon to reinforce their perceived expertise. By shaping their identities in line with societal expectations of authority, scammers seek to establish trust while facilitating compliance from unaware victims.

However, the façade constructed by scammers is inherently fragile and becomes increasingly evident when challenged. The pressure to maintain these personas leads many scammers to experience cognitive dissonance, especially when faced with the absurdity and resistance posed by scambaiters. As their identities come under scrutiny, scammers can exhibit defensive or aggressive behaviours, revealing the underlying insecurity that lies beneath their authoritative front. This emotional response underscores how the scambaiting interactions serve to not only expose the fraud but also reveal the complexities and vulnerabilities inherent in the scammers' constructed realities.

Conversely, scambaiters navigate a spectrum of identities throughout their engagements with scammers, demonstrating a remarkable ability to perform and adapt. Initially, they may portray themselves as naïve victims, adopting exaggerated confusion to draw scammers into their narrative. By fostering this initial relatability, scambaiters create an environment conducive to manipulation, undermining the scammer's attempt at authority before staging a tactical shift toward a more confrontational role. This transformation showcases their ability to reclaim agency, challenging the scammers' narrative and dismantling the illusion of control both parties have in these interactions.

Additionally, scambaiters often leverage collective community engagement during these exchanges. The live audience plays a vital role in identity construction, providing real-time moral support and suggestions that influence the scambaiter's approach. This dynamic reinforces a

sense of camaraderie and collective identity, further amplifying the scambaiter's persona as both protector and educator within the digital anti-fraud community.

Audience participation allows scambaiters to seamlessly shift roles, whether embodying a knowledgeable investigator challenging fraudulent behaviours or engaging with the audience as an advocate for victims' rights. Through the oscillation between different identities, victim, investigator, educator, the scambaiter not only critiques the scammers' authority but also shapes the public's understanding of fraud and victimhood. This complex interplay enhances their role as champions against scams, emphasizing the significance of unfettered identity negotiation during interactions.

The performances of both parties illuminate a broader dialogue about the nature of fraud, trust, and the societal dynamics surrounding the narratives of exploitation and resistance. Ultimately, the construction and negotiation of identities within scambaiting interactions encapsulate a rich tapestry of psychological engagement and social commentary. As scambaiters and scammers engage in these multifaceted exchanges, they enable a critical reflection on authority, victimhood, and the broader implications of deception in a digital world, thereby contributing to a nuanced understanding of consumer protection and vigilantism.

4.6 Emotional Dynamics and Psychological Influence

The emotional exchanges that occur between scammers and scambaiters during their interactions reveal a multi-layered dynamic of manipulation, resistance, and absurdity that influences the trajectory of each engagement. Both parties utilise a wide range of emotional expressions, although their intentions and impacts differ significantly, resulting in complex interplays characterized by tension, humour, and psychological manoeuvring.

Scamming operations revolve around a calculated exploitation of emotions aimed at influencing and manipulating victims towards following the instructions and directions of the scammers. Contact is generally initiated under the pretence of authority, seeing scammers employ emotional expressions designed to evoke immediate fear and urgency within their targets. Common phrases that reference accounts being compromised or threats of legal repercussions serve to instil panic and to force victims into compliance while preventing them from critically assessing the situation and conducting the necessary due diligence. The emotional manipulations exerted by scammers create an overwhelming sense of urgency, aiming to compel individuals to react quickly, therefore enhancing the effectiveness of the scam at subduing rational thought.

This initial interaction can be described using a dramaturgical framework, where the scammer is an actor performing on a "front stage." This performance is a conscious effort at impression management, designed to project a specific "front" of authority and legitimacy. The scammer's script, tone of voice, and use of technical jargon are the props and costumes used to define the situation for the audience, the victim, and create a convincing portrayal of a helpful agent.

Moreover, the narratives crafted by scammers intertwine urgency with empathy, falsely presenting themselves as helpful figures trying to assist the victim. Scammers often seek to establish levels of emotional connections with their targets, which is integral to their strategy that aims to lower defences and create rapport. Towards this end, scammers frequently iterate that their main goal is to help and support their target, and they can even go as far as to try to create an association to familial relationships through phrases such as, for example, “You’re like a grandparent to me”. Such expressions facilitate not only trust but also amplify emotional dependency as victims, blurring the lines between trust and exploitation.

The emotional dependency is further solidified through scammers attempting to prevent the victim from disclosing their current circumstances so as to reduce the likelihood of an external party intervening and revealing the fraudulent nature of the interaction. This is often attempted through seeding distrust between the victim and official figures that could potentially disrupt the scam, or by crafting scenarios in which the disclosure of specific information would put those who become involved at risk. For example, in instances where victims are directed to physically access their banks in order to transfer money, the scammer might have crafted a narrative where the bank staff is not to be trusted and would instruct the victim to use specific wording in order to avoid revealing critical information.

Similarly, in narratives that involve the threat of legal action, the victim might be advised that disclosing any of the information made available to them could put their loved ones at risk and as such, it is safer for the victim to not disclose anything to anyone. Such strategies not only reduce the likelihood of third parties becoming aware of the scam and intervening but also serve to isolate the victim, further increasing the emotional dependency on the benevolent scammer. The ability to project confidence and authority while masking deceptive intentions is therefore a key element in the scammer’s arsenal.

As the interaction progresses, however, the emotional landscape can shift dramatically, particularly when faced with the challenges presented by scambaiters. Faced with confrontation or non-compliance, scammers may reveal underlying vulnerability or frustration, as evident in their defensive responses or aggressive shifts in tone when their narratives are being poked at and deconstructed. Such emotional volatility not only signifies a breach of the

controlling narrative but also exposes the fraudster's reliance on manipulation, rather than genuine rapport.

Similarly to scammers, scambaiters also leverage a complex repertoire of emotional expressions that serve to challenge and expose the scammer's intentions. For scambaiters, this often involves the use of humour, absurdity, and mockery – strategies that serve to empower them and redirect the emotional engagement of the conversation. By employing comedic elements, scambaiters are able to craft an environment that not only entertains but also disarms the emotional manipulation that is frequently wielded by scammers. The emotional responses that scambaiters employ in their interactions with scammers – frequently featuring sarcasm, exaggerated confusion, and playful defiance – contribute to the flipping of the script of authority, positioning the scambaiter as a savvy participant rather than a passive victim.

The scambaiter's actions represent a dramaturgical counter-performance, intentionally designed to disrupt the scammer's presentation of self. The scambaiter engages in "breaking frame," a process where the established definition of the situation is actively challenged and undermined. By refusing to play the assigned role of the compliant victim, the scambaiter introduces interruptions that expose the artificiality of the scammer's front. This deliberate refusal to cooperate in the performance reveals the scammer's script as a fabrication and forces a confrontation with the underlying reality of the interaction.

The play on emotional dynamics challenges the scammer's perceived authority, enabling the scambaiter to reclaim agency and disrupt the oppressive atmosphere often associated with scamming scenarios. Another significant aspect of scambaiters' emotional expression can be found in their ability to navigate the emotional toll of their interactions in a balanced manner. Despite the underlying absurdity, balancing between humour and the weight of manipulation is vital. The acknowledgement of the emotional distress caused by scammers upon their victims is mirrored back in the scambaiter's responses, evoking empathy while simultaneously using those emotional responses as part of their strategy to resist the scam attempt. Although scambaiters may identify with the frustration experienced by potential victims, their emotional awareness enables them to leverage these frustrations effectively, thus transforming what would be a discouraging interaction into a comedic performance of resilience.

Audience engagement further complicates the emotional landscape of scambaiting interactions. As viewers use comments, reactions, and other mechanisms of interaction, they amplify the emotional expressions of both the scambaiters and the scammer but are also able to give rise to emotional responses of their own. The collective humour that is shared among viewers promotes a sense of community and solidarity in opposing fraud, enabling scambaiters to benefit from audience support as they navigate the scammers' manipulations. From a

dramaturgical perspective, the audience becomes part of a "performance team" with the scambaiter, cooperating to maintain a shared definition of the situation where the scammer is the object of ridicule.

Furthermore, community participation can bolster the scambaiters' confidence, enabling them to engage more persistently with the scammer. The expression of emotional responses by the audience creates a feedback loop that influences the scambaiter's approach in real-time, with positive reinforcement from the audience being encouraging, while frustration or disappointment may prompt the scambaiter to adjust their approach. While this feedback mechanism can be useful by acting as a level of accountability for the scambaiter, it can also be detrimental. For example, where audiences might become particularly frustrated, they might start calling upon the scambaiter to engage in increasingly hostile and aggressive activities against scammers. In such instances, the scambaiter also acts as a quasi-moderator, seeking to address the audience engagement in a manner that is constructive and that acknowledges their contribution, while not allowing it to hijack the scambaiting activities, potentially turning them into non-constructive retaliatory practices.

The dynamics of prolonged engagements in scambaiting interactions carry profound psychological implications for both scammers and scambaiters, influenced by the interplay of power, identity, and emotional manipulation. The encounters that emerge often unfold as an amalgam of psychological phenomena, wherein the roles of victim and perpetrator morph, challenge, and redefine themselves throughout the process.

For scammers, prolonged engagements with scambaiters can induce significant cognitive and emotional strain. As the scambaiter employs humour, absurdity, and challenges to the scammers' authority, the scammer's carefully constructed identity often begins to unravel. This process is akin to forcing the scammer's "backstage" self into public view. The frustration, anxiety, and anger that emerge are expressions of this backstage reality, revealing a "break in character" when the actor can no longer maintain their role. This forced transparency exposes the conflict between the competent persona they are trying to project and their inability to control the interaction, leading to a breakdown in their constructed persona.

As a result of this subversion of expected roles, scammers may end up suffering from psychological conflict. The initial positions held by scammers, underscored by their strategic use of authoritative language and deceptive scripts, quickly shift as the scambaiter asserts control through laughter, mockery, and playful non-compliance. This is frequently observed in the form of anxiety, frustration, and even anger, reflecting a breakdown in their constructed persona. Additionally, as interactions become longer and as more and more of the scammers' attempts at defrauding their victim fail, they become increasingly likely to attempt to shift their

strategies by pivoting into a different type of scam, potentially introducing new characters and imposing a new narrative. Interactions that have started as typical refund scams have been observed to evolve into romance scams, social security scams, and even attempts at identity theft, with scammers having reached points where they were unable to pursue their initial strategy, but not having reached a point where they can give up on their target. Such transitions are often done clumsily and add another layer of complexity for scammers who now have additional details and narratives to enact while attempting to maintain at least a semblance of consistency with previously established details.

The tensions that emerge from the role reversals that occur within scambaiting interactions give rise to a state of cognitive dissonance in scammers, where their beliefs or self-perception are contradicted by the reality of their actions and the actions of others. For example, scammers who may equate their dishonest occupation with skill and cleverness may experience internal conflict when confronted with a scambaiter's dismissive ridicule, failure to comply, or exposure of tactics. The internal struggle that emerges can lead to feelings of inadequacy, resentment, or even defensiveness, further complicating the emotional landscape of the engagement. In some instances, scammers might even resort to aggressive responses to mask their insecurity and reassert a sense of power. This psychological discomfort can linger well beyond the interaction, potentially leading to reflective changes in self-perception.

Commonly, scammers' defensive strategies are most visible when scambaiters confront the scammer, revealing that the identity of the victim has been a façade all along. In such scenarios, scammers often make claims that they knew all along about the interaction being a "prank", seeking to salvage a sense of control, as well as seeking to paint themselves as superior to the scambaiter confronting them by switching the topic to the financial lucrativeness of scams as opposed to scambaiting, or by attempting to paint scambaiting as a futile endeavour through boasting about the number of victims that continue to become defrauded by such schemes. Also very commonly, once the revelation takes place, scammers have been observed to employ abusive and offensive language aimed at the scambaiter, becoming particularly confrontational, but also relatively dismissive of the entire interaction due to the newfound intel that there was no real victim to begin with.

The transformation observed in scammers at the point of revelation, while amusing through their realisation of having been duped, also contains an element of resolution to the cognitive dissonance. Throughout the entire interaction, scammers are being frustrated by their failure to achieve their objectives and by the inability to reconcile this failure with their self-perception of competent fraudsters. However, the revelation that the entire interaction was crafted by a

scambaiter provides a resolution to this conflict, enabling scammers to dismiss the interaction and move on relatively easily.

In light of this, some scambaiters have started moving away from revealing and confronting scammers, seeking to maintain the façade for as long as possible. This movement has enabled scambaiters to not only achieve lengthier interactions (some of which span over multiple months) but also to reveal how scammers would occasionally revisit previous targets to make new attempts at extracting funds from them. When the scambaiters would reveal their identities and engage in confronting scammers, this would frequently be the end of the interaction with scammers blocking all contact and moving on. However, in the absence of revealing, even when the interaction is over due to scammers having given up, they are still left to reconcile what has transpired and are occasionally observed to return to the target after a certain amount of time has passed. This creates a longer-lasting emotional impact from having been scambaited with the resolution of the revelation no longer being available leaving scammers to confront their own incompetence at having failed to achieve their goals.

The environment within which scammers operate is also worth noting, with many scamming operations being organised as high-pressure environments where targets are being set and where performance is key. The “employment” in such organised call centres may breed a culture of competition that intensifies levels of stress, fostering an atmosphere where failure to achieve objectives can evoke feelings of shame and inadequacy. Thus, prolonged interactions with scambaiters can exacerbate such pressures through creating situations where scammers not only face scrutiny from the scambaiter, but also contend with the consequences of not meeting the expectations of their peers and “employer”. The interactions captured by scambaiters frequently observe scammers interacting with their “colleagues” brainstorming ideas to progress the scam or expressing frustration when their attempts fail. This teamwork element adds another layer of emotional complexity where, on the one hand scammers benefit from the support of their peers, while on the other hand failures are more readily evident to these same peers, opening the scammer up for criticism and mockery from their “colleagues”.

Scambaiters also encounter psychological ramifications from prolonged engagements with scammers. While scambaiters generally have a greater degree of control over the interactions, and they position themselves as empowered leaders in the fight against fraud, the continuous exposure to the manipulation and deceit employed by scammers can take an emotional toll. The scambaiters’ experience is generally characterized by a range of emotions including humour, frustration, empathy, and even anger at times, reflecting the complexities of navigating a system built on exploitation. The demand to maintain a façade of confidence and mastery

while engaging with scammers requires emotional labour, which can contribute to psychological fatigue over time.

This is especially visible in instances where scammers use scripts that are particularly unpalatable. For example, one relatively common scamming script revolves around informing the victim that one of their loved ones (generally a child or a grandchild) has been involved in an accident and the target now needs to provide payment for legal fees to assist their loved ones in navigating the consequences. In one example, a scammer has been observed to include additional details adding to the psychological weight of the scenario, claiming that the other party to the accident was a pregnant lady who is now in critical condition. Such scenarios, due to the heavy emotional manipulation that is involved, might be difficult to navigate for scambaiters as they need to remain a convincing victim in order to most effectively string the scammer along and disrupt their operations.

The repeated need to confront scammers who employ aggressive and manipulative tactics can also evoke feelings of compassion fatigue or vicarious trauma. Engaging with scammers who target vulnerable individuals, such as the elderly, may lead scambaiters to feel an emotional tether to these potential victims. This connection can amplify the emotional burden felt by scambaiters, particularly if the scambaiter's interactions highlight the harsh realities faced by those victimized by scams. This can be amplified even further in instances where scambaiters have had personal experiences of a loved one being defrauded, as some scambaiters have on occasion disclosed to their audiences. As scambaiters perform their roles, the distress experienced by victims may resonate deeply, giving rise to feelings of indignation, empathy, and even a sense of responsibility to protect others. Such feelings could also become more potent in instances where scambaiters interact directly with real victims in attempts to prevent them from being defrauded or to provide support and resources in the aftermath of a scam.

The introduction of humour as a coping strategy can serve to offset the psychological strain. Reframing the situations experienced through comedic interactions allows scambaiters to gradually cultivate resilience, utilising humour to alleviate some of the psychological weight of confronting manipulation and deceit. Humour also better enables scambaiters to reclaim a narrative of empowerment, transforming the potentially distressing engagements into opportunities for entertainment and learning. However, the reliance on humour can also mask deeper emotional responses that, if unaddressed, may lead to cumulative psychological distress over time. This is where stepping away from the humour and entertainment aspect of scambaiting can be helpful, allowing scambaiters to address the issues of online fraud with the seriousness warranted, thus avoiding the suppression of the unpleasant side of scambaiting.

The communal aspect of scambaiting also plays an important role in the emotional dynamics that come into play. With viewers actively participating in the narrative, scambaiters benefit from a source of reinforcement, as well as a feedback system which enables them to more effectively navigate the interactions. Viewer engagement fosters a sense of solidarity and support, mitigating feelings of isolation that could stem from the emotional labour involved in scambaiting. However, audiences could also present an external pressure for the scambaiter who might be seeking to maintain audience engagement and to keep the interactions entertaining. This adds a layer of complexity to the emotional landscape within which scambaiters operate, with the pressure to live up to audience expectations having the potential of triggering self-doubt or heightened stress.

The intricate dynamics that emerge in the scambaiting engagements observed reveal a complex web of psychological consequences that manifest for both scammers and scambaiters. Navigating through a system characterized by manipulation, authority, and community has an impact upon both parties, creating a continual evolution of self-perception shaped by their respective experiences and reactions. The interplay of power, identity, and emotional responses underscores the broader societal implications of these exchanges, emphasizing the need for continued awareness, education, and support systems to address the challenges inherent in combatting online fraud.

4.7 Chapter Summary

This chapter has provided a detailed analysis of the intricate and adversarial relationship between online scammers and scambaiters. It establishes that the conflict is characterised by a strategic symmetry, wherein scambaiters "fight fire with fire" by mirroring the deceptive tactics of their opponents. The analysis demonstrates how scammers employ a sophisticated arsenal of emotional manipulation, social engineering, and technological tools to create a façade of authority. In response, scambaiters adopt and invert these very strategies, using elaborate deception not for financial gain but for resource attrition, intelligence gathering, and operational disruption.

Central to the chapter's analysis is the application of Erving Goffman's dramaturgical framework. The interactions are deconstructed as complex theatrical performances, with scammers and scambaiters acting as performers on a "front stage." The chapter examines how both parties use technological "props" (such as remote access tools and voice modulators),

carefully constructed "scripts," and meticulous "impression management" to control the narrative and deceive their respective audiences.

A key contribution of this chapter is the introduction of a critical distinction between two primary modes of scambaiting: investigative and engagement-focused. It is argued that the latter, with its emphasis on long-form, humorous, and performance-based content, represents the clearest manifestation of entrepreneurial digilantism. This form is directly shaped by the demands of online content platforms, where the need to create entertaining and monetisable media influences the very tactics used for disruption.

The analysis further delves into the construction of identity, contrasting the fragile, authority-based persona of the scammer with the fluid, multifaceted identity of the scambaiter, who shifts between roles such as naïve victim, knowledgeable challenger, and public educator. Finally, the chapter explores the profound emotional and psychological dynamics of these prolonged engagements. It details the cognitive dissonance and frustration experienced by scammers as their performances are systematically dismantled, and it considers the emotional labour and potential for burnout faced by scambaiters. Through this multi-faceted examination, the chapter reveals scambaiting as a complex interplay of performance, psychological warfare, and identity negotiation, inextricably linked to the modern attention economy.

Chapter 5 Discussion: The Social Implications and Future Stages of Scambaiting

5.1 Introduction

This chapter delves into the intricacies of scambaiting, interpreting the findings from Chapter 4 to unpack the central thesis of this research: that scambaiting has evolved into a form of entrepreneurial digilantism. By drawing from the qualitative data and employing theoretical frameworks, particularly the dramaturgical analysis of Erving Goffman, this discussion explores the multidimensional aspects of scambaiting. It will analyse how the roles, power dynamics, and performances described in the previous chapter serve a dual purpose of both vigilantism and content creation.

The chapter will examine the symbolic language, humour, and community engagement that characterize the scambaiting phenomenon, linking them to the central argument about how monetisation and audience engagement reshape vigilante practices. Central to this discussion are the identities constructed within the scambaiting ecosystem and the strategies employed by scambaiters to manage and expand their operations.

By exploring the interplay between the findings presented in Chapter 4 and the theoretical concepts of performance, impression management, and community, this chapter aims to provide a nuanced perspective on the complex dynamics that underpin scambaiting. It will offer insights into how the content produced by scambaiters provides entertainment, education, and a combative measure against online scams, ultimately discussing the broader implications for society's approach to fraud in the digital age.

The 'Vigilantism' in Entrepreneurial Digilantism

In the digital age, where online fraud presents a ubiquitous threat, scambaiting has emerged as a captivating blend of vigilantism, entertainment, and education. This practice engages individuals who take it upon themselves to expose and undermine the deceptive strategies employed by scammers, often in humorous and absurdist ways. This section will explore the vigilante dimension of this phenomenon, outlining the arc of scambaiter practice across seven key areas. It begins by examining the core motivations that drive scambaiters and how their methods compare to traditional anti-fraud strategies. It then delves into how scambaiting

functions as a form of public accountability, followed by a critical look at the ethical ramifications of their actions and the specific moral considerations they face. The discussion then turns to the scambaiters' own self-perceptions and concludes by analysing how their construction of victimhood impacts broader community understanding.

Scambaiters not only aim to protect vulnerable populations from exploitation but also strive to instil a sense of accountability among both scammers and potential victims. Through their creative tactics, they cultivate narratives that highlight the importance of consumer awareness and resilience in the face of digital deception. However, the motivations and ethical considerations surrounding scambaiting reveal a complex interplay between altruism and the potential for harm, inviting a critical examination of its role within the broader spectrum of online crime prevention and digital activism.

Scambaiting motivations and narratives

Scambaiting, as a form of digital vigilantism, emerges as a multifaceted response to the endemic issue of online fraud, driven by a confluence of motivations that underscore its dynamic nature. These motivations can be linked to theories of social control and anomie, where scambaiters act as informal agents of control in a digital space where formal mechanisms are perceived as failing, responding to a state of normlessness in online interactions. Central to the ethos of scambaiters is a profound impulse to safeguard vulnerable populations from financial exploitation (DEYOCLUB, 2021a). This motivation is frequently articulated by scambaiters themselves. For instance, the creator Scammer Payback explicitly states, "In my streams and videos I try to equip you guys with information on how to protect yourself and your family" (Scammer Payback, 2021a, 08:20). Similarly, Kitboga reflects on a pivotal moment where his work transitioned from a hobby to a mission: "An old woman who had been scammed for 6 years called in our system and we were able to walk her through... What started off as a joke has become very very real" (Kitboga, 2021a, 20:45).

Many scambaiters perceive their engagement as a moral imperative, stepping into the roles of protectors against the predatory machinations of scammers. This is often expressed in moments of emotional gravity, such as when Scammer Payback, after successfully intervening in a live scam, remarks, "This is what it's all about" (Scammer Payback, 2021b, 06:17), later reflecting that for scammers, victims 'are just numbers to them' (Scammer Payback, 2021b, 11:53). This self-conception informs and shapes the narratives surrounding their activities, framing scambaiting not only as a necessity but as a socially responsible endeavour aimed at fostering general awareness and resilience in the face of deception. The narratives constructed

within scambaiting often position scambaiters as champions of justice, embodying a sense of civic duty that motivates their proactive measures against fraudulent practices.

This performance of the moral protector can be analysed through Erving Goffman's dramaturgical framework, where the scambaiter is an actor on a "front stage", their digital platform, managing impressions for an audience. The scambaiter's "personal front" includes their adopted persona, tone of voice, and stated motivations, all of which are carefully curated to project an image of sincerity and altruism. For the performance to be successful, the audience must believe in the authenticity of the scambaiter's commitment to justice. Goffman distinguishes between cynical and sincere performers; in this context, scambaiters present themselves as sincere, convinced by their own act. Their expressions of anger or relief are not merely for show but are integral parts of a performance that validates their role as genuine protectors, thereby legitimizing their vigilantism in the eyes of their followers.

By publicly exposing scams and educating potential victims about the operational tactics employed by scammers, scambaiters strive to disrupt scam operations while simultaneously imparting critical knowledge to their audiences. This dual focus on exposure and education establishes a protective narrative where the intention is not merely to thwart scams but to cultivate a platform from which viewers can also assume roles of vigilance and advocacy. Integral to this narrative framework is the emotional resonance brought forth by the scambaiters' interactions with both scams and scammers. By highlighting the often dire consequences victims experience as a result of such fraudulent schemes, scambaiters reinforce their identity as moral custodians. This was evident when Scammer Payback expressed his anger upon learning a victim had lost $24,000, stating, "Dang it! Freaking pisses me off" (Scammer Payback, 2021c, 05:08). This element of emotional engagement intensifies their commitment to educating the audience, often through meticulously constructed content that demystifies the mechanics of diverse scamming techniques.

However, the vigilantism inherent in scambaiting prompts a deeper exploration of accountability within its framework. Through their activities, scambaiters challenge the prevailing anonymity and perceived authority typically associated with scammers. They use strategies that include humour, ridicule, and strategic questioning to dismantle the scripts scammers employ for manipulation. A clear example of this is seen when Kitboga derails a scammer’s script by asking with complete sincerity, "Do you ever put eggs in your coffee? ... just put a raw egg in there, a little mayonnaise..." (Kitboga, 2017, 05:05), an absurd question designed to break the scammer's control. This confrontation transcends individual scams, constructing a broader critique of a system that often permits such fraud to proliferate unchecked.

Furthermore, the narratives crafted by scambaiters emphasize a communal aspect that reinforces their self-appointed roles as watchdogs against fraud. The interaction between scambaiters and their audiences cultivates a collective approach to combating online scams, positioning viewers not merely as passive receptors of information, but as active participants in a community advocating for accountability and vigilance. This collective approach is evident in the language used by creators, such as Scammer Payback, who often uses the pronoun 'we' when describing interventions, thereby including the audience in the action, as seen when he notes, "We were watching a scammer's computer… and we saw a live scam ongoing" (Scammer Payback, 2022, 00:58).

Within the dramaturgical model, the audience is not a passive observer but an essential component of the performance itself. They function as a "team" in Goffman's sense, working together with the primary performer, the scambaiter, to sustain a shared definition of the situation. Through live chat comments, donations, and social media engagement, the audience provides real-time feedback that can shape the direction of the interaction. They affirm the scambaiter's moral stance, celebrate their victories, and collectively condemn the scammer. This interactive dynamic solidifies the legitimacy of the scambaiting enterprise, transforming it from a solo act into a collective ritual of justice. The audience, therefore, co-constructs the performance, validating the scambaiter's role and reinforcing the community's shared values against fraud.

In sum, the motivations and narratives surrounding scambaiting reflect an intricate interplay of altruistic intentions and ethical quandaries. Scambaiters engage in a nuanced performance that combines education, disruption, and accountability, all while negotiating the complexities of morality within their vigilantism. Their actions embody a difficult balance: while they strive to protect vulnerable individuals from exploitation, they must also contend with the potential ramifications of their methods, ensuring that their pursuits do not devolve into unsanctioned cruelty or desensitization to human suffering. This narrative complexity speaks to the broader implications of scambaiting within the digital landscape, highlighting the need for ongoing reflection about its role in the fight against online crime and the ethical considerations that accompany such forms of digital activism.

The Carnival of Scambaiting

The findings presented in Chapter 4 indicate that scambaiting cannot be adequately characterised as a purely functional act of disruption aimed at neutralising fraudulent activity. Rather, it emerges as an expressive, highly stylised practice in which play, humour, and

performance are constitutive, not incidental, elements. While Goffman's (1959) dramaturgical framework offers a valuable account of the structural composition of these performances, illuminating the roles, scripts, and stages through which the interaction unfolds, the cultural criminology concepts introduced in Chapter 2 deepen our understanding of their affective power and enduring appeal. Through this lens, the scambaiting encounter is reframed as an example of what Presdee (2000) terms the carnival of crime, a bounded, ritualised spectacle in which the normative rules of social interaction are inverted, authority is lampooned, and the pleasures of transgression are enacted both for participants and for an observing public.

Within this digital carnival, the scammer enters under the assumption of narrative control, occupying a position of apparent authority as the orchestrator of deception. Yet the dramaturgical logic of the carnival dictates a reversal: the ostensible antagonist is re-scripted as the fool, their authority dismantled through absurdist provocation. The scambaiter assumes the role of the carnivalesque trickster, a figure whose power lies in subverting expectations, destabilising social roles, and introducing incongruous elements that erode the scammer's composure. The query by Kitboga, "Do you ever put eggs in your coffee? … just put a raw egg in there, a little mayonnaise…" (Kitboga 2017, 05:05), is emblematic of this tactic. On a pragmatic level, the question serves no functional purpose in dismantling the logistical infrastructure of the scam. Instead, it operates as a deliberate frame-break (Goffman 1974), disrupting the scammer's performance and introducing a moment of absurdity designed to amuse the audience and elicit visible frustration from the target. Here, the humour is not incidental; it is the vehicle through which symbolic domination is achieved, transforming the scammer from manipulator to manipulated.

The pleasure experienced by both the audience and the scambaiter in witnessing this inversion cannot be dismissed as trivial amusement. While it does not amount to sadism in a clinical sense, it aligns closely with the affective dynamics that cultural criminology identifies as central to transgressive practices, specifically, the pursuit of edgework (Lyng 1990), where actors seek the adrenaline and affective intensity of playing at the limits of normative boundaries. In the context of scambaiting, this edgework is not expressed through physical danger but through performative confrontation in which the scammer's confidence is systematically eroded, and their narrative control is reappropriated for comic and moral effect. The laughter of the audience, like that of the carnival crowd, becomes a form of symbolic violence, compounding the scammer's humiliation while reinforcing the scambaiter's position as orchestrator of the spectacle.

Further, the operational environment of scambaiting corresponds closely to Williams's (2007) notion of the potential space of crime, a liminal zone in which reality and fantasy are

interwoven, enabling transgression to be enacted in a creatively bounded and controlled environment. The technological apparatus, virtual machines, voice changers, spoofed telephone numbers, and fictitious identities, functions as both protective barrier and theatrical set. Within this hybridised space, the scambaiter is liberated from real-world vulnerability, able to "play" with the scammer through extended improvisation, deceptive role-play, and the deliberate cultivation of narrative absurdity. This aligns with Bakhtin's (1984) understanding of the carnivalesque as a site of licensed misrule, a temporary suspension of hierarchies in which the ordinary order is displaced by a parodic counter-order. The destruction of the scammer's operation thus occurs not through direct confrontation but through a performative, humorous process that entertains even as it undermines.

Importantly, this carnivalesque dimension is not peripheral but foundational to the conceptualisation of scambaiting as a form of entrepreneurial digilantism. The entertainment value of the performance is integral to its sustainability, particularly where monetisation through streaming platforms, advertising revenue, or audience donations is concerned. The humour, the absurdity, and the playful humiliation of the scammer constitute the "core product" that engages audiences, builds parasocial relationships, and ensures repeat viewership. In this context, the scambaiter operates not merely as a digital vigilante but as a carnival ringmaster, managing the tempo, tone, and dramaturgy of the encounter in order to maximise affective impact and viewer gratification. The role is therefore dual: the scambaiter is both moral entrepreneur, pursuing a public good through disruption of criminal activity, and performer-impresario, orchestrating a commercial spectacle of transgression for a paying public.

Seen through this synthesis of dramaturgical analysis, cultural criminology, and theorizations of the carnivalesque, scambaiting emerges as a deeply mediated, affectively charged cultural practice in which moral purpose and playful subversion are inseparable. It exemplifies the proposition that acts of deviance and control in the digital age are as much about performance, audience engagement, and symbolic inversion as they are about instrumental objectives. The laughter of the crowd is not an afterthought; it is the sound of a carnival in full swing.

Comparison to traditional counter-fraud strategies

When compared to other counter-fraud strategies, scambaiting presents a rather unique and engaging grassroots approach, contrasting with more traditional methods promoted by government agencies, financial institutions, and cybersecurity organizations. There are distinct

contributions that scambaiting brings to the broader fight against fraud which both contrast with and complement traditional approaches.

From a methodology perspective, scambaiting engages directly with scammers by using humour, absurdity, and role-playing to expose the deceptive practices employed by perpetrators of online fraud. Frequently, scambaiters construct a specific "personal front" as part of their performance on the "front stage" of the YouTube stream, creating interactive narratives that captivate both scammers and audiences. This approach is characterized by real-time interactions, where the scambaiters disrupt the scammer’s agenda while simultaneously providing education and entertainment.

Applying Erving Goffman's dramaturgical framework, the scambaiter's actions can be understood as a deliberate "performance." In this context, scambaiters construct a convincing "front" to define the situation for their audience, the scammer. This front involves managing a "personal front," which includes elements like voice, mannerisms, and tone, as well as the "setting," such as the use of virtual machines and voice modulators to create a believable, albeit fabricated, reality. The goal of this impression management is to present a crafted version of oneself to influence the scammer's perception, making the fabricated identity appear authentic and credible.

Through the integration of this playfulness within scambaiting strategies, scambaiters are able to effectively highlight the absurdity of the scripts and scenarios used by scammers. This is evident in the data where scambaiters use absurd questions or scenarios to break the scammer's script and control, a tactic frequently employed by creators like Kitboga to both entertain and expose the scam's illogical nature (Kitboga, 2017).

Using Goffman's framework, the dynamic between the scambaiter (performer) and the scammer (audience) is central to the interaction. The scambaiter's performance is meticulously crafted for two distinct audiences: the primary audience (the scammer) and the secondary, wider audience (e.g., YouTube viewers). For the scammer, the performance is designed to manipulate their perception of the situation, leading them to believe the fabricated persona is real. Simultaneously, the performance is tailored for the viewing audience, who are "in on the secret" and derive entertainment and education from watching the deception unfold. This dual performance requires immense "dramaturgical discipline," as the scambaiter must maintain their character and control the flow of information to prevent the performance from being discredited. Any break in character could reveal the deception to the scammer and collapse the entire interaction.

It needs to be noted however that, in the content produced, different scambaiters have different approaches which vary in how much of these playful interactions are captured and presented, as well as how much humour is employed. Some of the content is more educational in nature, focusing on explaining how scammers operate, rather than on the showcasing of interactions that take place. Nevertheless, the presence of humour and role-playing is still felt, although in varying degrees depending on the piece of content observed.

This variation in approach can be explored through Goffman's distinction between the "cynical" and "sincere" performer. A sincere performer genuinely believes in the reality of their own act, whereas a cynical performer is not taken in by their own routine and has ulterior motives. Scambaiters often exhibit a hybrid of these motivations. On one hand, they can be seen as sincere in their stated goal of fighting fraud and protecting victims, a belief that fuels their efforts. On the other hand, the performative nature of their content, the focus on entertainment, and the potential for online fame and monetisation suggest a cynical dimension, where the performance is a means to other ends, such as audience growth and revenue. This duality does not necessarily imply a contradiction; rather, scambaiters may be sincerely committed to their anti-fraud mission while cynically employing dramaturgical techniques to achieve it for the "good of the community."

In contrast, traditional anti-fraud strategies often involve more structured campaigns and the use of educational materials that are disseminated by regulatory bodies, financial institutions, and law enforcement agencies. These methods usually rely on passive engagement via public service announcements, brochures, and information websites that aim to educate consumers about recognizing and reporting scams. The onus is therefore placed upon the consumers becoming educated and vigilant to resist scamming attempts, which partly overlaps with the educational element of scambaiting, but lacks its immediacy, creativity, and effectiveness. These traditional strategies are often limited in how much information can be communicated, as well as in the medium of communication, whereas scambaiting benefits from a more flexible and more extensive educational experience by capturing and dissecting real-life scamming occurrences.

From a law enforcement perspective, the combatting of online fraud through formal investigations and prosecutions is becoming less and less effective due to the increasingly global aspect of scamming operations, and to the significant volume which outmatches the resources available to law enforcement strategies. Law enforcement agencies are increasingly struggling to combat the escalating global cyber-fraud threat. The traditional legal principle of territoriality often fails in multi-jurisdictional cyber contexts (Territoriality, 2025), and international bodies like INTERPOL remain severely under-resourced, with just over a thousand

staff facing millions of cross-border fraud cases (Button, 2025). Local agencies are overwhelmed, some manage around a thousand cybercrime calls per month and acknowledge that even doubling personnel would not suffice (Moloney et al., 2022). In the UK, only about 4 percent of fraud reports via Action Fraud result in prosecution (Button, 2021), while elsewhere, massive scams like pig-butchering are orchestrated by global syndicates whose operations now span Southeast Asia to Africa and beyond (UNODC, 2025). Even when law enforcement achieves successes in recovering stolen assets, as in the U.S., these efforts recover only a minuscule fraction of the sums lost to fraud, illuminating a profound mismatch between criminal efficiency and enforcement capacity (Solomon, 2024).

Additionally, formal law enforcement bodies are restricted by rules and regulations which dictate what they can and cannot do so as to avoid potential abuses of power. Scambaiting on the other hand does not suffer from the same restrictions and, through the investigative work being conducted, scambaiters have often been observed to engage in collaboration with law enforcement by collecting and submitting identifiable information about scammers in the hopes that this will enable formal law enforcement bodies to more effectively take action.

This is substantiated by data where scambaiters explicitly state their cooperation with authorities, such as Scammer Payback noting, "We got the address and we were able to give it to law enforcement" (Scammer Payback, 2021a, 13:17). Other long-form scambaits are explicitly undertaken to compile evidence, such as bank account details and personal information, for reporting to the relevant authorities (Kitboga, 2021b; Jim Browning, 2022b) In this regard, the effectiveness of the scambaiters' investigative activities can be beneficial to law enforcement endeavours to combat online fraud in instances where the latter has the necessary scope and resources to take action on the intelligence produced through scambaiting, creating a relatively collaborative relationship.

However, some scambaiters have expressed dissatisfaction with the effectiveness of law enforcement efforts in combating online fraud, with Scammer Payback remarking on his frustration with "traditional security and especially how consumers aren't really protected very well" (Scammer Payback, 2021d, 08:23). Some of the content produced even hints at how it often seems that corruption within local law enforcement actually generates a layer of protection for scamming operations who are able to use bribes and social connections to avoid having their operations dismantled.

5.1.1 Scambaiting as a form of public accountability

Scambaiting, as a unique form of digital vigilantism, manifests through a profound commitment to protecting vulnerable populations from the predatory practices of scammers. Scambaiters see themselves as moral agents who engage in proactive measures to expose fraudulent activities, fostering a culture of accountability that compels scammers to confront the consequences of their actions. Their self-perception as guardians against exploitation is rooted in a narrative that emphasizes social responsibility and a duty to inform and educate potential victims.

This moral stance creates a compelling framework through which scambaiters justify their actions, positioning their endeavours as a necessary counterbalance to the impunity often enjoyed by scammers operating in the shadows of the digital landscape. The underpinning motivations of scambaiters reflect a strong desire to safeguard those who are susceptible to financial exploitation. This protective narrative is often articulated through the detailed documentation of scams within their content, where scambaiters systematically dismantle fraudulent claims and reveal the manipulative strategies employed by scammers.

By doing so, they serve not only to inform audiences about the tactics used to deceive them but also to undermine the façade of legitimacy that scammers strive to project. The act of publicly confronting scammers empowers scambaiters, shifting the balance of power in these interactions and allowing them to reclaim agency from those who typically wield authority through manipulation.

Additionally, scambaiting serves as a mechanism for public accountability, illuminating the vulnerabilities inherent in scamming narratives. By exposing scammers in real-time through humorous and confrontational engagements, scambaiters diminish the social acceptance of these deceptive practices and cultivate a societal backlash against fraud. The playful yet assertive nature of scambaiting creates a culture in which scammers are not only met with ridicule but are also forced to reckon with the consequences of their fraudulent actions in a public forum.

This element of communal scrutiny enhances the moral narrative surrounding scambaiters' interventions, reinforcing the idea that accountability is not solely the responsibility of formal institutions but can be enacted by individuals operating outside traditional legal frameworks.

However, the ethical ramifications of scambaiters' actions merit careful consideration. While their intentions may be altruistic, scambaiting's confrontational nature can occasionally result in public shaming of scammers, potentially exacerbating the negative consequences faced by individuals who might already be ensnared in exploitative circumstances. Scambaiters must

navigate this ethical terrain with vigilance, ensuring that their pursuit of justice does not devolve into cruelty. By maintaining a focus on education and empowerment, scambaiters can preserve their moral high ground, striving to uphold the dignity of all individuals involved while emphasizing accountability and consumer protection.

In conclusion, scambaiting operates within a vigilantism framework that empowers individuals to combat fraud and hold perpetrators accountable. With a focus on public exposure and education, scambaiters embody the role of guardians, challenging the normative acceptability of scammers' tactics. While ethical challenges persist, the overarching narrative of vigilantism underscores the importance of community action in the fight against exploitation, fostering an environment where fraud is increasingly met with resistance and public accountability is championed by individual actors.

5.1.2 Ethical ramifications

Scambaiting can be understood as a form of digital vigilantism, where individuals proactively engage with scammers to expose fraudulent practices, ultimately protecting potential victims. Central to this practice is the self-perception of scambaiters as guardians of vulnerable populations, framing their actions through narratives that emphasize social responsibility and public accountability. By interrupting scams and arming viewers with knowledge about deceitful tactics, scambaiters position themselves as moral actors in a digital landscape fraught with exploitation.

This self-positioning can be analysed through a dramaturgical lens, as proposed by Erving Goffman. The scambaiter’s public-facing persona operates on a "front stage," where they perform the role of a moral guardian for an online audience. This performance is carefully constructed to convey competence, righteousness, and control, thereby managing the impression given to viewers. The scambaiter, as the actor, engages in a form of "expressive control," meticulously crafting their interactions with the scammer to align with the heroic narrative they wish to project. The authenticity of this performance, however, is a key consideration. While the stated goal is victim protection, the performance itself is also dependent on generating engaging content, suggesting a potential tension between the altruistic role being performed and the practical requirements of maintaining an audience.

The ethical ramifications of scambaiters' actions, however, are complex and multifaceted. While scambaiters often frame their activities as altruistic guardianship, there exists a tension between their self-defined roles and the potential harm they may inadvertently inflict. This

dichotomy becomes particularly pronounced when considering the emotional impact on scammers, who, despite their manipulative practices, may experience feelings of humiliation and distress during confrontations. The interactions may evolve into public spectacles where scammers are ridiculed, raising ethical concerns regarding the extent to which scambaiters should counter deceit with mockery. Such acts risk emphasizing punitive measures over restorative justice, challenging the notion of vigilantism as purely protective.

Moreover, the likelihood of desensitization among audiences is a pertinent concern. The entertainment value generated through humour and absurdity can overshadow the serious underlying issues related to fraud, potentially diverting attention from the very real consequences faced by victims of scamming. This can be analysed as a dynamic where the "working consensus" between the performer (scambaiter) and the performance team (audience) shifts to prioritise entertainment. As viewers engage with scambaiting videos, their reactions may reinforce a culture of ridicule rather than fostering empathy and understanding for those affected by fraud. This raises critical questions about the responsibilities that scambaiters hold in balancing entertainment with ethical considerations and the potential risks of perpetuating negative stereotypes about both scammers and victims.

Within Goffman's framework, the audience is not a passive recipient of the performance but an active participant that helps to co-construct its meaning. Viewers who engage with scambaiting content by liking, sharing, and commenting become part of the "performance team," validating the scambaiter's actions and reinforcing the frame of the interaction. When the audience response prioritises mockery and ridicule, it signals an acceptance of the performance as entertainment rather than as a purely ethical intervention. This dynamic can lock the scambaiter into a specific role, where deviating from the entertaining, punitive performance risks losing the audience's approval. The "working consensus" established between the performer and the audience thus solidifies a reality in which the humiliation of the scammer is a central and accepted feature of the spectacle, potentially overriding deeper ethical considerations.

Additionally, scambaiters come face to face with the character and emotional complexities of scammers themselves. While scambaiters may perceive scammers as deserving of public scrutiny due to their unethical behaviour, there is a wider conversation about the systemic conditions that lead individuals into such fraudulent activities. Scammers may be influenced by socio-economic factors, which complicates the moral high ground that scambaiters often claim. Consequently, the perceived need to hold scammers accountable must be carefully weighed against a nuanced understanding of their motivations and the potential for reform.

Goffman’s distinction between the "front stage" and "backstage" is also useful for considering the complex reality of the scammers themselves. The scambaiter’s interaction typically only

reveals the scammer's "front stage" performance, that of a deceitful operative. This is the persona they have adopted to carry out the fraud. However, this perspective ignores their "backstage" reality, which may be shaped by the socio-economic pressures and systemic conditions mentioned. By forcing the scammer into a public confrontation, the scambaiter effectively drags the performer's flawed front stage into the spotlight, while the backstage, the context, the personal history, the potential desperation, remains invisible. This curated exposure simplifies the scammer into a one-dimensional villain, making it easier for the scambaiter and the audience to justify punitive actions without engaging with the more challenging nuances of the scammer's situation.

In conclusion, scambaiting operates within a framework of vigilantism that emphasizes protection and exposure but must also navigate the ethical complexities of public engagement with fraud. Balancing the roles of educator and entertainer, scambaiters face the challenge of ensuring their actions contribute positively to the discourse surrounding online fraud without devolving into mockery or discrimination. This framework of vigilantism may serve as an important lens through which to evaluate the evolving dynamics of scambaiting, fostering the need for ongoing reflection into the ethical considerations that accompany this unique form of consumer protection.

5.1.3 Moral Considerations

Considering the strategies and tools that scambaiters use towards achieving their objectives, there are a number of moral considerations that need to be addressed. At its core, scambaiting oscillates between the noble pursuit of safeguarding vulnerable populations and the ethical dilemmas inherent in the employment of deception like that used by scammers. Furthermore, scambaiters often operate in legally grey areas, engaging in actions that may contravene established legal frameworks. The phrase “fighting fire with fire” encapsulates this ethical conundrum, where scambaiters navigate the terrain of honesty versus deception in the pursuit of their objectives.

While such tactics are effective at combating online fraud, they simultaneously risk perpetuating a cycle of deceit that undermines the overarching idea of moral integrity. Reliance on methodologies that mirror those of the scammers necessitates an ongoing critical reflection among scambaiting communities regarding the philosophical implications of their actions and the perceived legitimacy of employing similar means to achieve seemingly noble ends.

Applying Erving Goffman's dramaturgical framework, the scambaiter's engagement can be understood as a meticulously crafted performance. The scambaiter, as an actor, operates on a "front stage" when interacting with the scammer, employing a persona, complete with a fabricated backstory, specific mannerisms, and emotional responses, designed to be convincing. This is a conscious act of impression management, where the goal is to present a believable victim persona to manipulate the scammer. The "back stage" is the private space where the scambaiter can drop the act, plan their next moves, and perhaps confer with other members of the scambaiting community. This is where the performance is prepared and refined, away from the eyes of the "audience" (the scammer). The authenticity of the front-stage performance is paramount; any "break in character" could reveal the deception and end the interaction prematurely. Thus, scambaiters must skillfully manage the boundary between these two stages to maintain the illusion and achieve their objectives.

Moreover, the moral positioning observed within scambaiting communities tends to evoke a dichotomy in societal views – a tension that emerges from the contrast between perceptions of scambaiters as noble guardians of the online realm and the ethical ambiguity that results from their methodology. Supporters of scambaiting have been observed to praise its practitioners as contributing to a greater good, contributing to a heroic narrative being weaved around the practice. However, this framing coexists with critiques of the moral repercussions of ridicule and humiliation often directed at scammers during scambaiting interactions.

Further ethical complications come into play when considering the juxtaposition of scambaiters' entertaining and ethical conduct and the showcasing of emotionally charged responses from scammers who are frequently led into feelings of frustration and anger. The online community, in Goffman's terms, functions as a critical "audience" that shapes the scambaiter's performance. This audience is not passive; its reactions, feedback, and expectations actively influence the nature of the scambaiting drama. Scambaiters performing for a live or recorded audience may be motivated to escalate the drama, employ more elaborate deceptions, or provoke more extreme reactions from scammers to satisfy audience demand for entertainment. The audience, therefore, becomes a co-producer of the interaction, rewarding performances that are humorous, clever, or result in a satisfying sense of retributive justice. This dynamic can shift the focus from purely disrupting scams to creating compelling content, blurring the lines between vigilantism and entertainment.

The scambaiting practice also brings a level of disruption to traditional notions of authority and morality, positioning scambaiters as self-appointed agents of justice, while at the same time critiquing the shortcomings and inadequacies of law enforcement in tackling online fraud. While there is a noticeable level of cooperation that scambaiters engage in, mainly via reporting

actionable intelligence and collecting and providing evidence of wrongdoing, the perspective generally taken upon law enforcement is that it is not adequately dealing with the issue of online scams.

From a dramaturgical perspective, the motivation of scambaiters can be analysed through Goffman's distinction between "sincere" and "cynical" performers. A sincere performer genuinely believes in the reality they are presenting, while a cynical performer is not taken in by their own act and has ulterior motives. Many scambaiters may begin as sincere performers, genuinely motivated by a desire to protect victims and fight injustice. However, the performative nature of scambaiting, especially when conducted for an audience, can introduce cynical elements. The need to entertain, gain followers, or even generate income can become ends in themselves, potentially overshadowing the initial altruistic goals. It is plausible that many scambaiters exist on a spectrum between these two poles, genuinely believing in the good of their actions while also being aware of and playing to the entertainment value of their performance.

The subversion of authority raises broader societal questions around the legitimacy of taking justice into one’s own hands, particularly in instances where scams might emerge out of systematic socio-economic disparities that impact scammers themselves. A question thus emerges seeking to reconcile scambaiters’ accountability to their moral compasses with the environment within which they engage that invites the questioning of the institutional structures designed to uphold justice.

Ultimately, the moral landscape surrounding scambaiting is complex and necessitates a nuanced understanding of the interplay between tactics, intentions, and public perceptions. Scambaiters need to find an appropriate balance between entertainment, education, and effectiveness, juggling numerous ethical issues, while engaging within a framework where accountability is minimal.

5.1.4 Scambaiters’ self-perceptions

The realm of scambaiting presents a varied landscape of self-perception within which the participants who engage in deliberate activities have their identities shaped. The multiplicity of roles, expanding into the roles of educator, guardian, crime-fighter, and entertainer, reflects a complex interplay of motivations, ethical considerations, and community dynamics that highlight the broader ambitions of combatting online fraud.

At the heart of scambaiting, through the act of creating and making content available, there is a profound sense of moral responsibility manifesting in the scambaiters' self-perception as educators. This identity is firmly rooted in a commitment to disseminating knowledge that empowers individuals to navigate the complex and ever-evolving landscape of online fraud. The ambition to provide education and knowledge can be found in juxtaposition to the ambiguity and obfuscation that are often employed by online scammers to achieve their goals.

Scambaiters often frame their engagements with scammers as instructional opportunities, intent on elucidating the mechanics that are at work within various fraudulent schemes. Through documenting their interactions and disseminating them to wide audiences, scambaiters are able to transform their experiences into educational content that hopes to build resilience and to empower individuals to resist online scamming. Scambaiters often articulate their motivation as being an inherent duty to safeguard victims through education, with a significant part of the content created explicitly aiming to inform viewers so that they are equipped to resist manipulation.

This sense of advocating for consumer protection is underscored by the meticulous documentation and presentation of interactions with scammers, transforming this engagement into accessible learning materials. By breaking down the tactics employed by scammers, scambaiters strive to create a more informed audience capable of recognizing the vulnerabilities preyed on by scammers, thus becoming more resilient against them. The narrative woven by scambaiters also emphasizes a commitment to public accountability. Through their engagement with scammers, there is an ambition to dismantle the façade of legitimacy that fraudsters often construct and to challenge the manufactured authority that scammers project, thus revealing their true intentions. The process of revealing the deceptive nature of scams equips audiences and potential victims with insights that foster a broader understanding of how attempts at fraud can be identified and resisted.

However, scambaiters' self-perception is not limited to the role of educators. Another apparent dimension within scambaiting activities is their positioning as moral actors shielding vulnerable populations from the threat of online fraud. This self-identity is often rooted in narratives that place emphasis on a duty to safeguard those who may be targeted by scammers. This extends to deconstructing complex fraudulent enterprises, such as fake employment scams that lure victims into elaborate financial schemes (Pleasant Green, 2022). Through their interactions, scambaiters often present a vision of social responsibility, actively seeking to disrupt scamming operations.

Scambaiting activities are frequently framed as acts of vigilance, seeking to expose scammers to public scrutiny and dismantling the facades of legitimacy that are being falsely projected.

This positions scambaiters as an additional layer of deterrence for scammers who, if they find themselves in the crosshairs of scambaiting activity, risk being publicly exposed and even legal action if law enforcement becomes involved. Furthermore, by being able to intervene in scams that are being actively carried out against real victims, scambaiters further reinforce the perception of protectors through stepping in and preventing victims from falling prey to scammers' tactics.

In alignment with the educator and protector roles, scambaiters also wield the role of crime-fighter. Significant aspects of scambaiting revolve around disrupting and exposing scamming operations, with the hope that the disruption is sufficient to prevent the operations from being re-established. Part of these efforts consist of collecting evidence and passing it to appropriate law enforcement organizations so that official action can be taken against the perpetrators. This active engagement with formal channels of combatting online fraud supports the narrative of crime-fighting that is weaved within the activities carried out by the scambaiting communities.

Another prominent aspect of scambaiters' self-perception is that of entertainers. Scambaiting content often offers some entertainment value, with scambaiters needing to design and curate it in such a manner that it engages audiences. In many cases, scambaiters adopt elements that hold dramatic or comedic value, crafting narratives that are able to not only educate, but also captivate and entertain viewers. This dimension of scambaiting, while distinct from the more serious roles played by scambaiters, plays a significant part in promoting and spreading the content to help reach wide audiences. The impact of the content produced is thus enhanced, through increasing the likelihood of online users coming into contact with it.

The wielding of these self-perceptions is, however, also affected by some of the ethical considerations touched upon earlier. While the content created serves as a valuable resource for educating online users and potential victims, it also inadvertently offers insights to scammers themselves, enabling them to refine their tactics in light of learning about how they are being combatted. The widespread availability of scambaiting content raises questions about the unintended consequences of making such material public, creating a need to balance educational transparency with the risk of inadvertently empowering the very criminals that are being thwarted.

In addition, while scambaiters frequently position themselves as protectors, their focus on safeguarding potential victims does not always extend to giving appropriate considerations to the circumstances of the scammers themselves. Many individuals engaged in fraudulent activities might be situated in environments marked by poverty, lack of opportunity, or even coercion with their participation in scams being potentially more complex than mere criminal intent. By not awarding this dimension as much attention as it might necessitate, scambaiters

risk overlooking broader socio-economic factors that contribute to the persistence of online fraud, potentially framing all scammers as morally reprehensible without acknowledging the structural issues that may underlie their involvement.

The crime-fighting aspect of scambaiting also raises some legal and ethical concerns. While scambaiters may aim to disrupt scams and facilitate law enforcement efforts, their actions sometimes operate in a grey area of legality, running the risk of promoting activities that go against applicable legal frameworks. The use of deception, entrapment, and the unauthorized access to computer systems could, through their being in conflict with legal frameworks, undermine the legitimacy of scambaiting. As a result, scambaiters must navigate a fine line between informal, unsanctioned crime-fighting, and the risk of exposing themselves to legal repercussions.

Finally, there are ethical elements to be considered from an entertainment point of view as well. Through the pursuit of humour and entertainment, scambaiters run the risk of overshadowing the educational and protective functions of their activities in the interest of driving engagement. The desire to create entertaining content could lead to exaggerated or prolonged interactions with scammers, where the primary goal shifts from disrupting fraudulent activities to generating entertaining material.

Additionally, the pursuit of entertainment also extends into the ways in which viewers engage with the content. Some audiences may be drawn to scambaiting videos not necessarily for educational value, but for the spectacle of seeing scammers outwitted or humiliated. This dynamic could create an incentive for scambaiters to prioritise more sensationalistic content seeking to maximize viewer engagement.

Scambaiters as entertainers also face ethical dilemmas around their content when considering the perspective that it can be seen as glorifying vigilantism. Through the narratives that generally run through scambaiting content where the scambaiters are heroes battling against online fraud, there is a risk that audiences could be encouraged to adopt the perspective that ends justify the means. This portrayal could foster a culture of online vigilantism, where individuals might feel empowered to take the law into their own hands without adequate consideration for legal or ethical boundaries.

To summarize, the roles adopted by scambaiters – educator, protector, crime-fighter, and entertainer – illustrate the multifaceted nature of their engagement with online fraud, as well as the complexity which needs to be navigated in their activities. A complex interplay of motivations, ethical dilemmas, and societal responsibilities can be observed which brings about contributions to combatting online fraud, but also practical, ethical, and even legal

ramifications that need to be considered. Ultimately, scambaiting represents a unique and evolving response to the growing threat of online fraud, one that continues to raise critical questions about the balance between vigilantism, education, law enforcement, and entertainment.

5.1.5 Implications of Identity Constructions for Community Understanding of Victimhood

The process of identity construction, especially as engaged in by scambaiters, also has important implications for the understanding of victimhood. With the most common role taken up by scambaiters being that of the victim, they engage in a performance designed to lure scammers into a false sense of security and accomplish the goals of their scambaiting strategy. This enactment, however, is not without consequence, as scambaiters bear an ethical responsibility towards how they portray this role.

From a dramaturgical perspective, the scambaiter's performance is a carefully managed "front" presented on a "front stage" where both the scammer and the viewing audience are present. The scambaiter, as a performer, must convincingly project a "definition of the situation" in which they are a vulnerable and gullible target. This requires meticulous impression management, where every word, hesitation, and emotional outburst is a calculated part of the act. The authenticity of this performance is paramount to deceiving the scammer, yet it is this very performance of authenticity that complicates the portrayal of genuine victimhood. The scambaiter operates as a "cynical" performer in Goffman's terms, one who does not believe their own act, but must do so with enough skill to be perceived as "sincere" by their target.

A common risk in this endeavour is that, through the use of humour and absurdity, scambaiters might portray a potentially derogatory version of a victim, which could be interpreted by audiences as offensive, thus being detrimental to the efforts of scambaiting. To mitigate this, it has been frequently observed for scambaiters to use short breaks from their interactions with the scammers, where they address the audience directly. These moments function as a shift from the "front stage" performance to a "backstage" region of communication. By breaking character, the performer steps outside the scripted interaction with the scammer to remind the audience that despite the humorous nature of the performance, what is transpiring remains a serious issue that needs to be treated as such.

This is a recurring behaviour in Kitboga's videos, where after particularly intense interactions, he often pauses to speak directly to his audience, explaining the real-world danger of the tactics

being used (Kitboga, 2021e). For example, a scambaiter might engage in an exaggerated response as a result of a scammer's threats, presenting feelings of distress and turmoil which the audience might perceive as humorous due to their foreknowledge of the fabricated scenario. In such instances, scambaiters might address that the feelings they are enacting, while fake in the observed scenario, are real for an actual victim. This backstage commentary reminds the audience that despite the entertainment, real people suffer as a consequence of the scams tackled, thereby carefully managing the audience's interpretation and preserving the underlying sincerity of the scambaiting project.

Another way in which scambaiters impact the perception of victimhood is through their interactions with the audience. It has been commonly observed, in light of the absurdity of some scams, for viewers to raise questions as to how someone could fall prey to such ridiculous ploys. This question, while potentially posed in good faith, implies that real victims are perhaps less capable. In reality, scammers prey not on intellectual capabilities but on emotional responses that subdue rationality, as well as on individuals who might be more vulnerable due to disabilities, age, or a lack of technical savvy.

Generally, scambaiters tackle such questions by seeking to dispel any intentional or unintentional victim blaming. They draw attention to the emotional elements of scamming strategies and to the fact that, while the fraud might be evident in the observed scenario, it may not be in other circumstances. In these moments, the scambaiter's role as protector expands beyond protecting potential victims from scams, into also protecting existing victims from stigma and the shame frequently encountered by those who have been defrauded. By engaging with scammers in real-time, scambaiters highlight the absurdity and intricacies of online fraud, fostering a critical lens through which audiences can view potential victims.

This exposure to scamming tactics is transformative, as it redefines victimhood by challenging the associated stigma. Viewers gain significant access to the interactions that generally occur between scammers and their victims, enabling them to become more familiar with the details of how scammers operate. This level of access helps educate audiences that victims of scams are not merely naïve or uneducated, but are subjected to sophisticated manipulative tactics that exploit trust and emotional vulnerabilities.

Furthermore, scambaiting reveals how overwhelming scammers' approaches can be. Scambaiters have noted how, occasionally, scammers would end up making hundreds of phone calls and sending numerous messages, creating continuous pressure upon their victim and seeking to maintain control. Additionally, due to victims' details being frequently sold or exchanged through online channels, once someone has been defrauded, the likelihood of further attempts increases. This is also reflected in scambaiting, where a scambaiter subtly

indicating a past victimisation energizes the current scammer, encouraging them to become increasingly committed. Scambaiting can therefore create an empathetic community perspective, providing a nuanced understanding of victimhood that emphasizes the complexities of online fraud.

Finally, scambaiting communities also act as a forum for victims and their loved ones to share their stories within an understanding social group. Comment sections on YouTube videos have sometimes been used by those who have previously suffered from online fraud to share their stories and to express gratitude. Frequently, such comments also give credit to the humorous nature of the interactions and the role that these play in exposing scammers and neutralizing their attempts at establishing authority, thus turning their own deceptive strategies against them. On rarer occasions, individuals who have suffered at the hands of online fraud approach scambaiters and their communities seeking advice and support. Having developed knowledge and expertise, these communities are often able to provide information, signposting to appropriate resources, and comfort to those who seek aid.

In summary, the construction and negotiation of identities within scambaiting interactions significantly influence community perceptions of victimhood, agency, and accountability, creating spaces in which the stigma of victimhood can be combatted. By leveraging performance, scambaiters expose the manipulative tactics of scammers and challenge societal stigma against victims, fostering a supportive environment that emphasizes collective action and awareness. This process of identity evolution not only impacts individual narratives but also cultivates a broader understanding of vulnerability and responsibility within the digital ecosystem.

Audience Engagement and Community Dynamics

In the complex and multifaceted world of scambaiting, the audience plays an indispensable role that transcends mere observation, transforming viewers into engaged participants who shape the course of interactions between scambaiters and scammers. This section delves into how audience involvement not only fosters a vibrant community but also reinforces emotional connections, peer accountability, and shared purpose among members. As scambaiters and their audiences co-navigate the complexities of online fraud, the dynamics of their collective engagement contribute to an enriched experience that combines entertainment, education, and ethical considerations.

However, this relationship is not without its challenges; the potential for audience disruption and ethical dilemmas underscores the need for a thoughtful approach to scambaiting,

balancing the pursuit of engagement with the underlying mission of protecting vulnerable individuals from scams. This deep level of audience engagement and community formation is not only a social phenomenon but also a crucial economic driver, creating the foundation upon which scambaiters build their monetisation strategies.

5.1.6 The role of the audience in scambaiting

The role of the audience in scambaiting is pivotal, enabling viewers to transition from passive observers into active participants who significantly influence the dynamics of the interactions between scambaiters and scammers. As the interactions unfold on platforms such as YouTube and Twitch, viewers are able to not only engage with the content but also participate in a collaborative effort to combat online fraud. This sense of involvement fosters a vibrant community where shared experiences amplify both the educational and entertainment aspects of scambaiting, reinforcing a collective identity dedicated to resisting scams.

Participation can take many forms, from providing financial contributions and support, to sharing the content across various platforms, and even to providing feedback and ideas for scambaiters to take into account. The type of participation available is usually dictated by the type of content, with livestreams enabling real-time contributions from the audiences, while pre-recorded and published content is more suitable for post-interaction discussions.

Real-time engagement, in particular, enables a highly interactive level of participation. Scambaiters often explicitly invite and engage with comments, suggestions, and reactions, creating a communal atmosphere in the fight against online scams. Involvement is frequently encouraged by inviting audiences to offer insights or strategies, either via comments or other mechanisms such as polls. Audiences might suggest specific lines of questioning, responses, or humorous scenarios for the scambaiter to explore. This influence can be direct and acknowledged, with creators crediting their live audience for providing a successful tactic, as one noted: 'That’s because of you chat. I want to put this on the record… if you hadn’t done that none of that would have happened…' (More Kitboga, 2022, 10:45:45). Aside from boosting engagement, such interactions also empower audiences, as they sense their contributions matter and make a direct impact on the unfolding narrative.

From a dramaturgical perspective, the scambaiter’s performance is a complex act of impression management conducted for two distinct audiences simultaneously: the scammer, who is an unwitting participant in the drama, and the viewing audience, who is in on the act. In this framework, the scambaiter’s front-stage performance, the persona, script, and actions

presented to the scammer, is meticulously crafted not only for the scammer's belief but also for the viewers' entertainment and edification. The awareness of this primary, knowing audience fundamentally shapes the performance; the scambaiter may incorporate asides, humorous exaggerations, or explicit explanations of their tactics that serve no purpose for deceiving the scammer but are crucial for engaging the viewers. Thus, while the performance must be convincing to the scammer, it is ultimately constructed for the audience.

The audience, in turn, plays a crucial role in validating this performance. Their real-time reactions, such as comments, likes, and shares, function as a form of social proof, affirming the authenticity and effectiveness of the scambaiter's act. This validation is not merely passive; it is an active co-construction of the event's success. Furthermore, the audience often provides essential backstage support. While the scambaiter is on the front stage, locked in a delicate interaction with the scammer, the live chat can function as a backstage crew, offering suggestions, providing information, and giving moral support. This collaborative effort blurs the line between performer and observer, making the audience an integral part of the dramaturgical team that works to sustain the performance and ensure its successful outcome.

The emotional investment of the audience is also important towards shaping the scambaiting experience. Shared laughter and camaraderie strengthen community ties as viewers are able to bond over the absurdities of scams and the strategic dismantling of the narratives pushed by scammers. As the emotional landscape evolves alongside the scenarios that unfold, viewers participate in experiencing amusement, frustration, and empathy, fostering a sense of belonging and shared purpose. This emotional connection reinforces the idea that scambaiting is not just an individual endeavour, but rather a collective community effort aimed at protecting the vulnerable and raising awareness about the dangers of online fraud.

Audiences also play a role in enhancing the educational aspect of scambaiting. By gaining insights into the tactics employed by scammers while witnessing real interactions captured by scambaiters, audiences become better equipped to not only tackle such situations were they to encounter them, but also to contribute to the dissemination of the educational elements that scambaiters impart. This makes the learning experience transformative in a sense, as it empowers viewers to emerge not only entertained but also better educated about the nuances of online fraud.

Moreover, the feedback loop created by audience engagement fosters a culture of accountability, with viewers playing a key role in ensuring scambaiters uphold ethical and moral standards in their interactions. The delineation of these standards is also co-created between the scambaiter and their audience, with positions being often discussed and negotiated, led by scambaiters and contributed to by members of the community. The discussion forums that

emerge within these communities enable the outlining of ethical positions, as well as collaborative explorations of the educational elements of scambaiting.

However, members of the audience do not always act or behave in the best interests of the wider scambaiting efforts. Scambaiters must always be aware of the risk of critical information being leaked, which enables disruptive viewers to jeopardise the scambaiting activities. Such risks create difficulties for scambaiters who, in addition to maintaining a certain level of protection from the scammers they interact with, also need to be wary of the dangers that their own audiences pose to their endeavours.

A common example of this is found in the frequent use of remote connection software. Such tools often require some sort of credentials to be communicated between the scammer and the scambaiter for the connection to be made. If the credentials are accidentally leaked to the audience, any viewer could similarly create a connection to the scambaiter's system, actively interfering with the efforts of establishing authenticity. Such interference is often significant enough to dissuade scammers from further engagement, serving as a warning that they might not be interacting with a legitimate victim and thus ruining the scambaiting efforts. Other sensitive information also consists of email addresses and telephone numbers; the leaking of contact details to the audience has on occasion resulted in viewers getting in touch with the scammer and warning them. The motivation for such disruption is interesting, and is often attributed to online trolling, but it will not be explored further in this piece of research as it is outside the scope.

Another interesting observation is the fact that scambaiting communities and channels are often leveraged by the very scammers they are combatting towards carrying out fraudulent schemes. Scambaiters have noted how, on rare occasions, nefarious actors leave comments in YouTube comment sections, masquerading as the respective scambaiter and seeking to promote a fraudulent scheme. The impersonation of popular content creators is not a rare occurrence, with social media platforms being riddled with fake accounts that are designed to look legitimate to lure unsuspecting victims into various scams. However, while this phenomenon is not isolated to scambaiting content, its presence within comments to scambaiting videos is particularly interesting since it targets, on the one hand, users that would be familiar with such fraudulent schemes, but on the other hand, users who might lack in familiarity and would be seeking to become better educated on the topic of online fraud. Scambaiters and their communities have been observed to be critical of the social media platforms that allow and enable such behaviour, since such scam attempts present a real risk to online users.

Finally, members of scambaiting communities have also been observed to leverage their artistic abilities to creating content based on more memorable scambaiting interactions or characters. Such content can take the form of images which are generally created to portray characters or specific interactions, or even songs which incorporate more memorable emotional outbursts that are triggered within scammers. Such contributions add to the entertainment value of the scambaiting endeavours and further strengthen the solidarity of communities through creating and disseminating art that has specific meaning for the community within which it has been created.

In summary, audiences represent a significant element of scambaiting efforts, and their engagement goes beyond simple passive content consumption into more active participatory activities. This participatory culture that is being cultivated is characterized by shared emotional experiences, educational growth, and collective responsibility. As scambaiters and viewers interact, they are able to co-create narratives that support the missions of scambaiting while fostering a sense of community solidarity. However, audiences also have the ability of impeding scambaiting efforts, placing the onus upon the scambaiter to ensure that the engagement is balanced so that audiences are able to participate without becoming detrimental to the scambaiting endeavours.

5.1.7 Emotional connections and solidarity within scambaiting communities

The emotional dimensions of audience engagement are particularly interesting, as audiences have been observed to not only echo and validate the emotions portrayed by scambaiters, but also to enhance their commitment to scambaiting through fostering emotional connections and a sense of community. Such emotional investment contributes to the sense of solidarity that can be observed within scambaiting communities, with scambaiters and users alike engaging in actively developing and enabling each other to better counteract the threat of online fraud.

From a dramaturgical perspective, the audience's comments and donations are a form of immediate feedback that validates the performance, assuring the actor that their "definition of the situation" has been accepted. This is crucial to the co-construction of the performance. The scambaiter acts as a performer on a digital "front stage," managing impressions for both the scammer and their audience. The audience, in this framework, is not a passive observer but an active participant that validates the scambaiter's "front", that of a competent, humorous, and morally righteous vigilante. Their real-time reactions serve as cues that affirm the performer's chosen script and encourage improvisational risk-taking, thereby solidifying the collective "definition of the situation" in which the scammer is a deserving target of deception and ridicule.

For example, collective laughter and shared amusement at absurd situations encourage scambaiters to push against the scammer's authority with humour and creativity. Furthermore, the emotional investment also cultivates a community that promotes an emphatic approach towards potential victims, as well as a collective feeling of indignation at the strategies employed by scammers. As viewers resonate with the scambaiters' purpose to protect the vulnerable, they engage in adopting an identity that aligns with advocacy against scams, potentially becoming motivated to engage beyond the digital arena. This could translate into sharing scambaiting content in their own networks or discussing strategies for recognizing and avoiding scams in personal conversations.

The sense of solidarity is also strengthened by shared experiences where participation in the achievement of certain milestones can be a uniting element for the community. Examples of such achievements include various records for how much of the time of various scammers was wasted, or the successful deployment of a new scambaiting strategy or tool. The sense of collective contributions towards such achievements becoming possible has been observed to resonate strongly with audiences who will often express encouragement and support for scambaiters upon the reaching of various milestones.

Another element commonly used towards enhancing community engagement and solidarity consists of various memes and cultural references. This also overlaps with how humour is occasionally achieved, but scambaiters (and community members) have been observed to reference cultural elements in their engagements which, through recognition from the community, enhance the sense of solidarity and strengthen the community identity around shared cultural norms.

Moreover, memorable interactions or particularly successful scambaiting attempts may become exalted to cultural symbols specific to the respective scambaiting community within which they emerge. For example, particularly humorous responses employed by either scammers or scambaiters, through their appeal to the community, might become commonly used by the scambaiter in subsequent interactions. This serves not only to reference the circumstances within which the response emerged, but also as a way to leverage a shared symbol towards acknowledging the community.

This is frequently done in a verbal fashion, but references to previous interactions are also sometimes included in a visual manner, with content creators being able to manipulate their backgrounds to include elements that add cultural and comedic value. For example, in reference to the frequency with which scammers may seek payment through gift cards, scambaiters might create a background that humorously simulates the gift card section of a store, with signage that references commonly used phrases by scammers.

Such visual elements are also employed towards enhancing audience engagement and fostering a sense of community. Scambaiters might also employ specific visual designs that reference and show gratitude to users for various contributions made to the scambaiting efforts. For example, through in-built functionality on Twitch or YouTube, scambaiters are able to set up visuals to automatically display when members of the audience make financial contributions to the channels, either as one-off donations or as subscriptions. Such contributions are also occasionally acknowledged verbally, although not consistently as this might often overlap with a pivotal point of the scambaiting interaction which the scambaiter might need to prioritise.

In sum, the transcending of audience engagement from passive consumption to active participation within scambaiting communities contributes to the cultivation of emotional connectedness and a strong sense of solidarity. This has positive contributions to the scambaiters' interactions, with scambaiters being able to rely on support and encouragement from their audiences, while also being beneficial towards the overarching missions of combating online fraud and disseminating educational messages to increase the resilience of consumers.

5.1.8 Implications of audience engagement for scambaiters' strategies

The sections above have touched upon some of the ways in which audience engagement is being promoted and identified how scambaiters' activities can benefit from the engagement being driven. In addition to the emotional support that scambaiters benefit from, their strategies and methods also benefit from the input provided by their communities.

While a significant part of the community may consist of online users who are seeking entertainment and education on the topic of online fraud, some members also bring professional expertise. Individuals with backgrounds in banking, cryptocurrency, or other relevant fields have been able to leverage their professional experience to support scambaiting endeavours. They provide domain-specific knowledge, thus filling in potential gaps in a scambaiter's repertoire. Furthermore, such members of the audience can also facilitate partnerships between scambaiters and other organizations, which can lead to the enhancement of both parties' ability to counteract online fraud.

Another critical area where audience engagement impacts scambaiting strategies is through the identification and reporting of scamming operations. The sheer volume of online fraudulent schemes makes it difficult for scambaiters and their teams to identify, research, and tackle them without audience support. Members of the audience who encounter scamming attempts

can submit these to scambaiters, alongside relevant information that enables them to take action.

For example, some scammers, attempting to avoid being scambaited, have introduced verification steps to ensure that the person contacting them is a legitimate victim. These steps might consist of verification codes included in fraudulent communications or questions about how the individual came across them. Audience members who report scams can include such details, enabling scambaiters to contravene these measures and establish authenticity in their approach.

Community members also contribute to the development and implementation of new scambaiting strategies by participating in brainstorming sessions or by providing advice and direction. Frequently, scambaiters engage in modifying their systems or developing tools that help disrupt scamming operations. While not always going into significant detail, scambaiters occasionally discuss such tools with their audiences, who can contribute ideas or suggest ways of implementing the necessary technology to achieve the desired effect. Such collaborations have enabled the creation of methods to frustrate scammers' attempts and increase visibility into their operations where they might have wished to obscure elements of their activity.

Audience engagement also introduces important layers of accountability within the scambaiting ecosystem. Viewers, acting as a collective conscience, can prompt scambaiters to maintain ethical standards. This encourages actions that prioritise education and victim protection over mere mockery or humiliation of scammers. This can lead to discussions about the ethical dimensions of scambaiting, as community members express their desire for the process to maintain focus on exposing fraud without devolving into unnecessary cruelty. Such engagement not only reflects broader societal values regarding ethics and responsibility but also indicates a community capacity for self-regulation, ensuring that the mission of scambaiting remains centered around supporting and protecting victims rather than simply entertaining audiences.

From a Goffmanian perspective, the role of the YouTube audience is paramount in shaping the scambaiter's performance. Erving Goffman's dramaturgical theory, which uses the metaphor of a theatrical performance to understand social interaction, is particularly applicable here. The scambaiter operates on a "front stage", the YouTube video or livestream, where their performance is visible to the audience. This front-stage behavior is a carefully constructed presentation of self, designed to meet the audience's expectations for what a scambaiter should be: clever, witty, and effective in disrupting scams. The awareness of being watched by this audience compels the scambaiter to manage their impression, ensuring their actions align with the heroic, technologically adept persona they project. The performance, therefore, is

arguably dual-focused; it is directed at the scammer as the immediate target of the interaction, but its ultimate purpose is for the consumption and validation of the YouTube audience.

The audience's role extends beyond passive observation; they are active participants who validate the success of the performance. Positive comments, shares, and financial support are forms of feedback that affirm the scambaiter's competence and the legitimacy of their "show." This validation is crucial for the scambaiter's motivation and the continued production of content. Furthermore, the audience provides essential "backstage" support. As Goffman describes, the backstage is a private region where the performer can relax, prepare, and drop the front. In this context, the scambaiting community functions as a backstage team. They provide technical advice, report new scams, and collaboratively brainstorm strategies away from the live performance with the scammer. This backstage collaboration is where the performance is refined and prepared, allowing the scambaiter to appear polished and effective on the front stage. The audience is thus not merely a group of spectators but an integral part of the dramaturgical team, co-constructing the performance and ensuring its continued success.

5.2 The 'Entrepreneurial' Engine: How Monetisation Reshapes Vigilante Practice

In the realm of scambaiting, monetisation functions as a vital engine driving the initiatives of creators who seek to expose and disrupt online fraud. As they blend entertainment with education, scambaiters not only strive to raise awareness about scams but also navigate various avenues of revenue generation that align with their mission.

This section delves into the intricacies of these monetisation strategies, illuminating how they shape the operational landscape of scambaiting while fostering a sustained commitment to community engagement and consumer protection. By analysing the diverse methods employed, from ad revenue and direct viewer support to merchandising and sponsorships, it becomes clear that the financial aspects of scambaiting are deeply entwined with its effectiveness and ethical considerations, highlighting both the advantages and challenges faced by those dedicated to combating digital deception.

5.2.1 Content monetisation in scambaiting

The monetisation strategies employed by scambaiters are integral to sustaining their operations, reflecting a complex interplay between content creation and audience engagement. In an increasingly digital landscape, scambaiters have diversified their approaches to generate

revenue while fostering community awareness around online scams. This section explores the various monetisation methods utilised, emphasizing their implications for the sustainability of scambaiting as a form of vigilantism.

One of the primary methods scambaiters leverage is the generation of ad revenue through platforms like YouTube and Twitch. By producing engaging and educational content that resonates with viewers, scambaiters can attract significant audience engagement. Scambaiters often encourage audiences to like, subscribe, and comment on their videos, which boosts visibility and, in turn, ad revenue potential. This feedback loop underscores how effective content engages viewers not only for entertainment but also for crucial awareness regarding scams.

However, the generation of ad revenue is contingent upon the content adhering to the terms and services of the platforms used. This creates challenges for scambaiting, as the monetisation of content is often limited, or downright impossible, due to the nature of the material portrayed. On YouTube, for instance, content can be demonetised for potential violations of harassment and bullying policies, the inclusion of harmful or dangerous acts, inappropriate language, and privacy concerns for those depicted. Scambaiting is prone to fall foul of such policies, especially content that seeks to unmask and reveal the identities of those involved in online fraud. Furthermore, many advertisers prefer to avoid placing their ads on content involving criminal or unethical behaviour, a category into which scambaiting could be placed despite its educational intent. Consequently, ad revenue is often available in a limited measure compared to other content hosted on the same platforms.

In addition to advertising income, scambaiters capitalise on direct viewer support through features like Super Chat on YouTube or donation functionalities on Twitch. These tools allow audience members to contribute financially during live streams, highlighting the emotional investment viewers have in the scambaiting process. By appealing to a sense of community protection and engagement, scambaiters create a non-traditional ecosystem where financial contributions reflect a shared commitment to combat fraud. Such income streams are somewhat more reliable as they primarily depend on the direct financial commitment of viewers.

From a dramaturgical perspective, these direct financial interactions during a live performance are crucial for maintaining the authenticity of the scambaiter's presented self. In Goffman's terms, the scambaiter is on the "front stage," performing the role of a digital vigilante for a live audience. The use of Super Chat and direct donations functions as a powerful form of audience feedback, akin to applause in a traditional theatre, which immediately validates the performance. This real-time financial affirmation does more than just provide funds; it signals to

both the performer and the wider audience that the portrayal of competence, wit, and moral righteousness is being successfully received. The audience, by contributing, moves from being passive spectators to active participants in the drama, co-constructing the legitimacy of the scambaiter's actions and reinforcing the collective belief that the performance is an authentic and worthy endeavour.

Crowdfunding platforms such as Patreon and Ko-fi also form a significant part of the monetisation landscape. By offering exclusive content, behind-the-scenes access, or personalized interactions, scambaiters cultivate a loyal base of supporters willing to fund their mission. This model fosters a direct and durable relationship between creator and supporter, enhancing engagement while emphasizing the communal effort against scams and further solidifying the bond within the community.

Merchandising emerges as another effective strategy, allowing scambaiters to create branded products that resonate with the community’s ethos of vigilance. The sale of merchandise not only provides financial resources but also fosters brand identity and unity among supporters. Items like clothing and accessories carrying anti-scam messaging or inside jokes from previous encounters serve as tangible expressions of commitment to a shared cause, enhancing community visibility and solidarity. This approach also allows scambaiters to further solidify cultural elements that emerge within their communities, as memes or particularly amusing moments from interactions with scammers can be adapted into merchandise, contributing to the financial sustainability of their operations.

Applying Goffman's framework, merchandise can be understood as the symbolic "equipment" or "props" that signify an individual's membership in a particular performance team. The scambaiter and their audience form a team collaboratively engaged in the performance of anti-scam vigilantism. By purchasing and displaying merchandise, audience members adopt the team's "collective representation," visibly aligning themselves with the scambaiter's front. This act extends their role beyond the digital space of a livestream or video, allowing them to carry the performance into their everyday lives. The shared symbols on the merchandise reinforce the boundary between their team and that of the "others" (the scammers), strengthening the in-group solidarity and the shared reality constructed through the scambaiter's narrative.

Sponsorships and partnerships with relevant organizations represent another critical element of monetisation. By collaborating with businesses that align with their mission, such as cybersecurity firms, scambaiters can enhance their credibility and financial backing. Sponsored content that offers genuine products, such as antivirus software, provides a dual benefit by promoting necessary digital safeguards while generating revenue through affiliate earnings. These partnerships sometimes go beyond mere financial arrangements; scambaiters

occasionally combine their efforts with commercial or non-profit organizations to strengthen their efforts against online fraud through knowledge exchange and mutual support. For example, a collaboration with a remote access software company might grant scambaiters a specifically designed channel to report illicit use of the software, enabling the company to act more swiftly based on the established trust.

Finally, some scambaiters leverage their accumulated expertise to develop and commercialize their own protective tools. One example is Seraph Secure, software designed to prevent the installation of remote access tools, identify and block malicious websites, and alert users to potentially fraudulent activity in real-time. The development and sale of such tools are intrinsically linked to the scambaiting activities; the expertise gained from encounters with scammers directly informs the system's design, while the software's commercial success creates an additional, sustainable revenue stream for the scambaiting operation.

The effective combination of these monetisation strategies illustrates how scambaiters have successfully integrated various channels to support their operations. As they navigate the landscape of fraud exposure, these creators are positioned as independent agents, creatively leveraging community engagement and technological platforms to foster awareness and resilience against online deception. The future of scambaiting hinges on these multifaceted approaches, impacting not only the sustainability of their efforts but also the larger discussion around consumer protection in the digital age. As scams continue to evolve, so too must the strategies that scambaiters employ to ensure they remain effective and relevant in their quest against online fraud.

5.2.2 The influence of monetisation on scambaiting strategies and ethical considerations

The monetisation strategies that have become available and are being employed by scambaiters play a central role in the development of their operations to more effectively combat online fraud. The ability to monetise their content enables scambaiters to pursue such activities in a manner that exceeds a simple hobby, transitioning into what could be considered the equivalent of full-time employment. The financial success achieved by some scambaiters has enabled them not only to focus on scambaiting as a full-time occupation but also to build teams around them to further support the combatting of online fraud.

Achieving financial self-sufficiency through scambaiting is beneficial in enabling scambaiters to dedicate more time and resources to these endeavours. However, it also has drawbacks in that

the livelihood of scambaiters in this position becomes relatively dependent on their scambaiting activities. This dependency introduces a critical tension that can be examined through a Goffmanian dramaturgical lens, particularly concerning the authenticity of the scambaiter's performance. Erving Goffman distinguishes between the 'sincere' performer, who believes in the impression fostered by their own performance, and the 'cynical' performer, who has no such belief and uses the performance as a means to another end. The introduction of monetisation risks shifting the scambaiter’s motivation from a sincere performance of vigilantism to a cynical performance of entertainment.

Initially, a scambaiter may be a 'sincere' performer, genuinely motivated by a desire to disrupt fraud and educate the public. In this frame, their actions are an authentic expression of their role as a digital crime fighter. However, as financial incentives become central to the sustainability of their work, the performer may become more 'cynical'. The primary motivation can subtly shift from the act of justice itself to the necessity of producing content that satisfies an audience and generates revenue. This does not mean the scambaiter no longer cares about fighting scams, but rather that the performance of fighting scams becomes a carefully managed presentation, tailored to maximize engagement and financial return. This shift fundamentally affects the credibility of their role; if the audience perceives the performance as driven by profit rather than principle, the scambaiter's legitimacy as an anti-fraud advocate is undermined, blurring the line between a social crusader and a content creator.

While audience engagement can enhance scambaiting strategies, it also introduces the risk of unintended negative consequences where scambaiters might become too focused on metrics. Scambaiters often aim to maximize their reach, both to spread educational content and to more effectively monetise their efforts. To this end, audience feedback provides valuable insights into which types of content are most likely to attract views and drive engagement. However, prioritising entertainment and audience appeal can lead to compromising the ethical standards that guide scambaiting.

When scambaiters focus too much on catering to the desires of the audience, they run the risk of veering into sensationalism or clickbait, potentially undermining their educational goals and weakening the broader fight against online fraud. For example, through observing that content capturing particular types of scams leads to more audience engagement, scambaiters might seek to repeat the production of similar content to continue reaching similar levels of engagement. This approach, however, disincentivises scambaiters from exploring other types of scams that might need investigating, thus reducing the potential educational impact they could achieve.

On a similar note, considering the importance of driving engagement for financial sustainability, scambaiters could succumb to methods of audience exploitation. This could translate into crafting their content in a manner that manipulates audience emotions, using outrage, fear, or schadenfreude to drive engagement. Such methods, while effective, could be perceived as disingenuous, shifting the focus from providing educational value toward exploiting viewers' emotions for clicks, views, or donations. Quite ironically, this would bring scambaiters in very close proximity to the scammers they are fighting against, with both parties engaging in some level of emotional manipulation towards achieving financial goals. While audience engagement is important, the attention given to driving it should be balanced with the wider efforts of combatting online fraud and educating users.

Scambaiters are required to navigate a complex environment that features numerous ethical and legal dilemmas, as well as questions around whether the ends justify the means. In such an environment, audience engagement is a potent element, but scambaiters need to ensure that they leverage it in a balanced and controlled manner to avoid the negative impacts it can threaten.

Conversely, having access to sufficient financial resources enables scambaiters to better invest in the development of new tools and strategies to combat online scams. This can lead to novel approaches to wasting scammers' time and disrupting their activities. Through creating new ways of frustrating scammers' plans, scambaiters not only become more effective but are also able to introduce a sustainable level of novelty in their content, thus contributing to audience retention and engagement. Furthermore, considering the rapid pace at which the online fraud environment evolves alongside technological advancements, scambaiters are in a sense required to keep up with such developments to maintain the necessary effectiveness to oppose the latest scams.

Access to financial resources is also beneficial from a content perspective, enabling scambaiters access to better tools to create engaging and compelling videos. Part of content creation consists of various processes relating to the editing and publishing of videos which can be labour-intensive and could be dependent on the availability of specialised software. Furthermore, there might also be an expectation from the audience that scambaiters create and make content available at a certain frequency, an expectation which if not met might result in a reduction of engagement. Financial resources are beneficial here not only for securing the necessary tools and software for video editing but also for outsourcing some content creation elements, enabling the scambaiter to dedicate more focus to perhaps more impactful activities against online fraud.

Finally, considering the complex legal and ethical landscapes being navigated by scambaiters, access to financial resources could also translate into access to professional guidance. Scambaiters have been observed indicating that they sought legal advice prior to engaging in specific strategies to ensure that their course of action does not carry them into legal trouble. Considering the complexities involved, access to such specialised advice is beneficial, not only in terms of keeping them out of legal trouble but also as an additional layer to keep scambaiting activities contained within an appropriate legal and ethical framework.

While the monetisation of scambaiting content brings about significant benefits, there are also concerns that need to be addressed. One of them, as mentioned, is the financial dependency that emerges between scambaiters and the content they create. This dependency could act as an incentive for scambaiters to shift focus towards driving engagement and audience growth at the expense of effectiveness and adherence to ethical standards. Furthermore, the financial success of some scambaiters could act as a model to others who might seek to engage in creating scambaiting content primarily to pursue financial goals. This is particularly problematic in instances where newcomers are less capable of navigating the complex ethical and legal landscape and might fail to strike an appropriate balance.

Another issue that arises is the dependency between scambaiters and the platforms they use. While multiple platforms are available for hosting content and enabling monetisation, a dependency is still established on the platform for maintaining streams of income. This can raise challenges when it comes to terms of service and the nature of the content being produced. For example, during the data gathering process, the researcher has come across YouTube videos that would later be taken down due to violations of terms of service, which also implies a loss of revenue. The restrictions that emerge can cause issues for some of the content produced by scambaiters, especially when it reveals specific companies or individuals involved in fraudulent schemes. This has even been weaponized by scammers who, upon being targeted, submit complaints through YouTube requesting the removal of content for privacy violations, thus creating a channel through which they can fight back and impact the financial element of scambaiting operations.

Shifting to more ethical considerations, the monetisation of scambaiting content presents a dilemma when the activities involve acting outside of prescribed legal frameworks. In instances where scambaiters might engage in illegal activities such as gaining unauthorized access to computer systems or unethical practices such as deception to obtain intelligence, the monetisation of the resulting content can be seen as problematic. The financial gain that results raises ethical questions and opens scambaiting to criticism. While scambaiting seeks to combat the use of illegal means to achieve financial objectives, it can sometimes appear to be

operating in a similar fashion. The main difference is that scambaiters are more transparent about their activities with their audiences, whereas scammers rely on deception for financial gain. Despite this, the ethical consideration remains.

An interesting dilemma also emerges when considering how funds are used by scambaiters to investigate and disrupt online scams. More advanced scams might require some form of “initial investment” to progress, otherwise the scammers are likely to disengage. In such scenarios, scambaiters could reach a point where the only way to proceed is to allow the scammers to receive that “initial investment.” The informed use of funds to gain this level of access could be seen as problematic since those funds turn into financial gain for the online scammers. Here, scambaiters must weigh the cost of access against the potential educational and disruptive benefits. It is worth noting that such approaches are similar to strategies employed by law enforcement, where a small gain is allowed to be made by those breaking the law in exchange for a greater benefit. The difference, however, is that law enforcement operates based on prescribed rules of conduct, while scambaiters must make these decisions within their communities and based on their own judgements.

Finally, it needs to be noted that scambaiting operations can differ greatly in the financial sustainability they have managed to achieve. Some are in a financial position to build a support team, while others might still be working towards generating sufficient revenue to become a full-time activity. At the same time, some scambaiters might be satisfied with engaging in scambaiting as a hobby or a part-time occupation. However, irrespective of their position, the implications of content monetisation remain relevant and worth considering to ensure that scambaiting is conducted in a manner consistent with the wider ambitions of combatting online scams and with adequate consideration for the complex ethical and legal landscape these ambitions entail.

5.3 Chapter Summary

This chapter synthesises the research findings to substantiate the central thesis that contemporary scambaiting has evolved into a form of entrepreneurial digilantism. Utilising Erving Goffman's dramaturgical analysis as its primary interpretive lens, the chapter deconstructs the multifaceted performances of scambaiters, examining how their roles, strategies, and interactions serve the dual purpose of unsanctioned justice and monetisable content creation. The discussion demonstrates that while scambaiting is rooted in the vigilante

motivation to protect the vulnerable, its modern practice is inseparable from its function as entertainment.

The analysis of the 'vigilantism' dimension reveals scambaiters' complex self-perceptions as educators, guardians, and crime-fighters who operate in a space where formal law enforcement is perceived as inadequate. Drawing on concepts from Cultural Criminology, the chapter argues that the practice is not merely functional but is also a "carnival of crime," where the humorous and absurd performance serves to invert power dynamics and ridicule the authority of the scammer for an engaged audience. This carnivalesque spectacle, it is argued, constitutes the core "product" of the scambaiting enterprise.

The 'entrepreneurial' engine of this phenomenon is then examined through an analysis of the diverse monetisation strategies, from ad revenue and direct donations to merchandising and sponsorships, that sustain these operations. The chapter posits that this economic imperative fundamentally reshapes the vigilante practice. This creates a critical tension, explored through Goffman's distinction between the "sincere" and "cynical" performer, where the need to generate engaging content risks prioritising entertainment value over the foundational mission of anti-fraud activism.

Throughout the discussion, the audience is positioned as a pivotal actor, functioning not as a passive observer but as an active "performance team" that co-constructs the narrative, provides resources, and reinforces community solidarity. Ultimately, this chapter concludes that scambaiting is a complex hybrid phenomenon: a legitimate form of grassroots digital activism that is simultaneously a commercial enterprise, whose methods, ethics, and very sustainability are inextricably shaped by the demands of the online attention economy.

Chapter 6 Conclusion

6.1 Future Trends in Scambaiting and Scamming

In considering the ever-evolving landscape of online fraud, it is clear that scambaiters will be continuously required to observe, understand, and act against new trends and strategies that emerge. Technological advancements will continue to play a significant role in the development of both online fraud and scambaiting strategies, as will any shifts in legal frameworks and measures taken by private corporations. The complexity of this landscape is unlikely to reduce, and scambaiters will need to continue adapting their strategies to remain effective.

A primary technological development that warrants close observation is Artificial Intelligence (AI), which poses a dual-edged threat and opportunity. For scammers, generative AI can create more convincing and personalized phishing emails, generate hyper-realistic deepfake videos for impersonation scams, and power sophisticated chatbots to engage victims in real-time, making such schemes harder to detect. This development will challenge scambaiters by raising the bar for authenticity and requiring more advanced technical skills to unmask AI-driven fraud.

Conversely, while recent advancements have reduced some barriers to entry for scammers, they have also equipped scambaiters with new, powerful tools. Scambaiters could use AI to analyse scam scripts at scale, identify patterns, and even deploy their own AI bots to engage and waste the time of scamming operations more efficiently. Further explorations of the potential brought by AI are necessary, especially considering that similar explorations are already taking place in the scamming world. Consequently, expanding the educational elements of scambaiting to encompass how AI is being used to perpetrate online scams will be essential to continue building community resilience against fraudulent online behaviour.

While scambaiting efforts have proven effective, a continued diversification of the scamming schemes being tackled would be beneficial. It is not necessarily true that future content will be predominantly centred upon tech support and refund scams; the data already shows a movement towards tackling other forms of fraud, and this is likely to expand. For instance, the exposure of complex financial fraud, such as the Ponzi schemes investigated by YouTubers like Danny De Hek, or the public confrontation of money mules by groups like Trilogy Media, who have confronted scammers in person (Trilogy Media, 2019), and whose archived work continues to document these real-world interventions (Trilogy Vault, 2021). As these more intricate scams gain prominence, scambaiting will likely evolve to include more financial investigation and real-world intervention.

Furthermore, the channel of engagement captured in most scambaiting content consists of phone interactions, yet a significant number of scams are perpetrated through social media or email. Recent content shows increased interest in tackling these other channels, and the continuation of this direction would contribute to a more comprehensive diversity of interactions becoming catalogued and disseminated.

As public awareness of scams grows, scammers may shift towards more subtle approaches that blend legitimate offers with fraudulent elements. This hybridization could manifest through operations that mimic established brands or organizations, creating a façade of authenticity. Consequently, scambaiters will need to refine their engagement tactics, focusing on critically assessing the legitimacy of scammer claims while employing humour and absurdity to expose their underlying methods.

The role of monetisation will also continue to shape future trends. As scambaiting becomes more popular and financially viable, there is a risk that the integrity of the practice could be compromised. The pressure to create entertaining content for revenue might overshadow the core mission of disruption and education. This could lead to a convergence with mainstream media formats, similar to television shows like To Catch a Predator or Scam Interceptors, where the spectacle of confrontation becomes the primary product. The continued popularisation of scambaiting, fuelled by monetisation, could therefore push the practice towards sensationalism, potentially impacting its credibility.

However, monetisation also enables scambaiters to develop more sophisticated tools and form professional collaborations. Recent partnerships between scambaiters and commercial entities, such as cybersecurity firms, have enhanced their operational capabilities and provided a model for future knowledge-exchange. Such collaborations influence the community aspect of scambaiting, contributing to its expansion and the wider promotion of the content created. The development of commercially distributed anti-scamming tools, born from scambaiter expertise, represents a significant market-based solution that could be further explored, offering potential benefits to both users and developers.

A final trend worth monitoring is the potential integration of scambaiting with formal volunteer programmes in law enforcement. Initiatives like the "cyber specials" in UK policing or the use of reformed cybercriminals by Dutch police suggest a move towards leveraging citizen expertise. Scambaiters, with their unique skills and intelligence-gathering capabilities, could theoretically serve as an informal intelligence arm for future policing efforts. However, such a move would be fraught with problems, including issues of accountability, the legality of their methods, the reliability of their intelligence, and the fundamental challenge of integrating unsanctioned vigilante activities into a structured, regulated system.

Stepping away from practical elements, the ethical and legal implications of scambaiting will continue to present a complex landscape for practitioners to navigate. The burden of ensuring a balanced approach to the multiple ethical and legal dilemmas inherent to the practice will persist and might even amplify as new methodologies and strategies are explored. To ensure the sustainable development of scambaiting, those who engage in such practices must give appropriate consideration to these ethical dimensions or risk detracting from the overarching mission of combatting online fraud.

The participation of the audience will also continue to play a critical role in the evolution of scambaiting. As viewers engage in real-time during live interactions, their contributions can enhance the scambaiter's strategic approach, allowing for vital adaptations based on audience sentiment. This collaborative dynamic between scambaiters and their viewers can cultivate a more informed and vigilant community.

In conclusion, the future of scambaiting will be defined by a continual interplay between evolving scam tactics and the adaptive strategies employed by scambaiters. Responsiveness to emerging digital fraud will continue to necessitate innovation in scambaiting tactics, with an emphasis on education, audience engagement, and humorous critique. As scams become more sophisticated, the goal of scambaiting will remain focused on empowering consumers, cultivating resilience against fraud, and fostering a collective commitment to protecting vulnerable populations.

6.2 Future Research

The phenomenon of scambaiting, as explored in this thesis, offers a unique and intricate intersection of online vigilantism, community engagement, and the evolving nature of digital fraud. This thesis has made a concerted effort to address its original problem statement: to explore the evolution of scambaiting and its role in combating online fraud through the lens of entrepreneurial digilantism. By analysing a substantial dataset of YouTube content, it has succeeded in identifying key strategies, identity constructions, and community dynamics that define modern scambaiting. While the researcher hopes that this work contributes to a better academic understanding of the scambaiting phenomenon, critical avenues remain available for future research which can further deepen understanding of the dynamics at play in this space.

A primary limitation of this study is its exclusive focus on content published on YouTube. As has been noted, scambaiters often utilise an ecosystem of online platforms to facilitate their operations. Future research would benefit from exploring scambaiting as it is captured outside of YouTube. Investigations of the interactions that emerge on platforms such as Twitch, Patreon,

and Reddit might provide additional insights not captured in the current research, thus expanding the existing body of academic knowledge on the multifaceted nature of this practice.

Furthermore, the vast majority of content analysed in this research was conducted in English. Future studies could extend beyond the Anglophone environments that dominate the current scambaiting discourse. Investigating scambaiting activities in non-English speaking contexts could illuminate crucial cultural differences that may influence tactics, motivations, and community responses. A focus on case studies from regions where specific types of scams are particularly prevalent could also contribute to a more comprehensive and globally relevant understanding of counter-fraud strategies and community engagement.

Another methodological limitation stems from the focus on video transcripts and the comment sections of scambaiting videos. While this qualitative data features multiple instances of scambaiters expressing their views, opinions, and ethical positions, future research could benefit greatly from more structured methods for exploring their perspectives. Semi-structured interviews, for example, would enable a more in-depth exploration of the ethical issues touched upon in this work, as well as the elements of identity and community that are central to the scambaiting ecosystem.

Such an approach would be particularly fruitful for applying a dramaturgical analysis to the scambaiter's craft. The YouTube video represents the polished "front stage" performance, where the scambaiter, as a performer, carefully manages their presentation of self for an audience. Interviews would grant access to the "backstage" region of their practice. Here, a researcher could explore the unglamorous preparatory work, the moments of frustration, the genuine emotional responses to the scammer, and the candid reflections on their motivations that are carefully filtered out of the public-facing performance. This would allow for a richer understanding of how scambaiters construct and maintain their vigilante identity, distinguishing between the performed persona and the individual behind the screen.

Considering the significant role played by audiences and communities, an ethnographic study also presents a compelling avenue for future research. By engaging directly with a scambaiting community through participant observation and interviews, a researcher could gain superior access to the cultural elements and tacit knowledge that may not be fully captured in the current body of work. This immersive approach could shed additional light on the use of cultural symbolism in scambaiting and the intricate community-building elements at play.

An ethnographic methodology is uniquely suited to observing the collaborative nature of scambaiting through the Goffmanian lens of "performance teams." Many scambaiting operations are not solo endeavours but rely on a team of individuals, including other

scambaiters, moderators, and active audience members, who collude to stage a convincing performance for the scammer. Ethnography would allow a researcher to observe the "dramaturgical discipline" required of this team: how they coordinate their actions, share information privately, and maintain a consistent front to successfully deceive the target and entertain their audience, thereby revealing the collective effort inherent in these complex digital interactions.

The technological aspects of scambaiting also warrant deeper exploration. Future research could focus not only on the sophisticated technological landscape navigated by scambaiters but also on the wider digital channels through which scammers operate. Adopting a cybersecurity-oriented approach to the study of scambaiting could provide technical insights that could prove influential in the evolution of scambaiting practices and broader anti-fraud strategies. Given the level of technological expertise often leveraged in these activities, such explorations would make a significant contribution to understanding the adaptability of scambaiting to the ever-changing landscape of online fraud.

Finally, further examination of the ethical and legal dimensions of scambaiting is crucial, particularly in exploring the line between acceptable vigilantism and potential overreach. Questions surrounding the moral obligations of scambaiters, especially in relation to their portrayal of both scammers and victims, merit deeper scrutiny. The implications of publicly exposing and shaming individuals who engage in online fraud, and the potential for misidentification or disproportionate consequences, require careful consideration. Such research could contribute to the development of ethical frameworks that scambaiters might adapt to ensure a more responsible and effective approach to countering online fraud.

In conclusion, while this thesis has made significant strides in understanding the phenomenon of scambaiting, the topic remains understudied and presents great potential for further inquiry. The avenues for future research outlined above have the potential to enrich the discourse surrounding digital vigilantism, enhance community engagement, and promote effective strategies for combating online fraud. As technological advancements and fraud tactics continue to evolve, so too will scambaiting. Ongoing research in this area is essential for fostering resilient digital communities empowered against online deception and exploitation, giving appropriate consideration to the complexity of the ethical and legal landscapes that accompany such efforts.

6.3 Final thoughts

In a world increasingly dictated by digital interactions, scambaiting emerges as a multifaceted response to online fraud. The examination of this phenomenon in its modern form reveals its complexity as an educational, entertaining, and ethically contentious practice. Through the dynamic interactions between scammers and scambaiters, an intricate clash of roles emerges, with each party negotiating power and identity within an ecosystem characterized by deception and resistance.

This interaction can be understood through Erving Goffman's dramaturgical framework, where the encounter between scambaiter and scammer is a carefully staged performance. The digital interface becomes the 'front stage' where each actor performs a role for the other, the scambaiter feigning vulnerability or greed, the scammer projecting authority or trustworthiness. Both engage in meticulous 'impression management' to make their assumed characters believable. The scambaiter's 'back stage' is their private operational space, the forums, chat groups, or recording setups where the performance is planned, scripts are written, and technologies are prepared, all hidden from the scammer's view. This backstage is where the authentic self, the scambaiter, rehearses the role they will play on the front stage.

Scambaiters, positioned as both vigilantes and educators, employ diverse methodologies and technological tools to expose and undermine scammers' fraudulent tactics. Their engagements serve not only to entertain and inform audiences but also to foster a collective consciousness around vigilance against scams.

From a dramaturgical perspective, the online audience is not a passive observer but an integral part of the performance. The audience validates the scambaiter's role, affirming their performance of competence and moral authority. Their real-time feedback, through comments, likes, and shares, can influence the scambaiter's script, encouraging certain tactics while discouraging others. This interactive dynamic solidifies the legitimacy of the scambaiter's actions, transforming a private act of deception into a public performance of justice. The audience, in essence, becomes a cast of supporting actors who, by bearing witness, help to frame the scambaiter's work as both authentic and socially valuable.

The implications of scambaiting extend beyond individual encounters, shaping broader societal perceptions of victimhood and agency. Through the emphasis placed on the emotional manipulation inherent in online scams, scambaiters are able to challenge prevailing notions of victim culpability, advocating instead for approaches driven by compassion and understanding of the complexities surrounding fraud.

The construction of the scambaiter's 'character' is central to this advocacy. To be effective, they must perform the role of an ethical and just vigilante, a character that requires constant and careful maintenance. This performance is a delicate balancing act; the scambaiter must appear clever enough to outwit the scammer but relatable enough to champion the average person. Their motivation, presented as a desire for justice and education, forms the core of this character. However, this performance is threatened by the very ethical ambiguities of their work. Actions that stray into excessive deception, humiliation, or sensationalism risk breaking character, potentially undermining the moral legitimacy they have worked to project to their audience.

The supportive communities that form around scambaiting efforts contribute to a culture of resilience and awareness, empowering individuals and communities to recognise and resist online scams. However, risks also exist in the blurred lines of ethical engagement, the need for constant negotiation of community norms, and the pitfalls of sensationalism, which necessitate ongoing reflection and evaluation of the practices explored.

Furthermore, monetisation strategies play a crucial role in sustaining these initiatives. Crowdfunding, merchandise, and advertising revenue allow scambaiters to pursue their work with the resources necessary to combat increasingly sophisticated scams, effectively funding the 'production' of their dramaturgical performances.

Ultimately, the ability to blend entertainment with education while navigating these significant ethical considerations will determine the trajectory of scambaiting in the evolving digital landscape. As the interplay between scammers and scambaiters continues to unfold, the commitment to consumer protection, community engagement, and the ethical ramifications of their actions will remain central to the scambaiting narrative.

In conclusion, scambaiting is more than just an attempt at disrupting online fraud. It encapsulates a vibrant community endeavour, highlighting the importance of collective action, societal reflection, and ongoing adaptability in the face of an ever-evolving technological landscape. The future of scambaiting promises further innovation and evolution, reinforcing the invaluable role of community-based approaches in the broader fight against deception in the digital realm.

List of References

Abrahams, R. (1987) 'Sungusungu: village vigilante groups in Tanzania', *African Affairs*, 86(343), pp. 179-196.

Abrahams, R. (2000) 'Vigilantism, state jurisdiction and community morality', in *Morals of legitimacy: Between agency and system*. Oxford: Berghahn Books, p. 107.

Action Fraud (2023) *National Fraud and Cyber Crime Statistics*. London: City of London Police

Adams, K., Hester, P., Bradley, J., Meyers, T. and Keating, C. (2014) 'Systems theory as the foundation for understanding systems', *Systems Engineering*, 17(1), pp. 112-123.

Adewunmi, O. (2024) 'Monetisation strategies for content creators', *IOSR Journal of Economics and Finance*, 15(6), pp. 57–66.

Adzimah-Alade, M., Akotia, C., Annor, F. and Quarshie, E. (2020) 'Vigilantism in Ghana: trends, victim characteristics, and reported reasons', *The Howard Journal of Crime and Justice*, 59(4), pp. 449-466.

Agnew, R. (2017) 'Revitalizing Merton: general strain theory', in *The origins of American criminology*. London: Routledge, pp. 137-158.

Agnew, R., Brezina, T., Wright, J. and Cullen, F. (2002) 'Strain, personality traits, and delinquency: extending general strain theory', *Criminology*, 40(1), pp. 43-72.

Ali, M.A. (2019) 'Consumer facing technology fraud: economics, attack vectors and mitigation', *Computers & Security*, 83, pp. 233-253.

Altheide, D.L. (2018) 'The media syndrome and reflexive mediation', in *Media logic(s) revisited: modelling the interplay between media institutions, media technology and societal change*. London: Palgrave Macmillan, pp. 11-39.

Anadi, S. (2017) 'Security and crime prevention in under-policed societies: the experiment of community vigilantism in Anambra State of Nigeria, West Africa', *Journal of Law, Policy and Globalization*, 60, pp. 122-129.

Anderson, K. (2024) 'The weaponization of trust: Emotional dynamics in social engineering', *Journal of Cybersecurity Psychology*, 12(1), pp. 45-62.

Anter, A. (2019) *The modern state and its monopoly on violence*. Colchester: ECPR Press.

Bajaj, P. and Edwards, M. (2023) *Automatic scam baiting using ChatGPT*. Available at: https://arxiv.org/abs/2309.01586 (Accessed: 15 August 2025).

Bakhtin, M. (1984) *Rabelais and His World*. Bloomington: Indiana University Press.

Bano, M., Zowghi, D. and Whittle, J. (2023) *Exploring qualitative research using LLMs*. Available at: https://arxiv.org/abs/2306.13298 (Accessed: 15 August 2025).

Basiru, A.S. and Osunkoya, O.A. (2019) 'Vigilante groups and policing in a democratizing Nigeria: navigating the context and issues', *Revista Brasileira de Estudos Africanos*, 4(8), pp. 128-146.

BBC (2021a) *How Kitboga uses AI to disrupt scam operations*. 19 March. Available at: https://www.bbc.co.uk/news/av/technology-56458267 (Accessed: 15 August 2025).

BBC (2021b) *Inside the scam call centres stealing millions*. 27 February. Available at: https://www.bbc.co.uk/news/av/stories-51660982 (Accessed: 15 August 2025).

Bencherki, N. (2017) 'Actor-network theory', in *The International Encyclopedia of Organizational Communication*. Hoboken, NJ: John Wiley & Sons, pp. 1-13.

Bernburg, J. (2019) 'Anomie theory', in *Oxford Research Encyclopedia of Criminology and Criminal Justice*. Oxford: Oxford University Press.

Bérney, M., Ondrus, J. and Holzer, A. (2024) 'Navigating the shadows of cyber vigilantism: a preliminary analysis of social dynamics and activities of scambaiting', in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*, Honolulu, HI, USA, 11–16 May, pp. 1–6.

Bichler, G. and Malm, A. (2015) 'The routine nature of transnational crime', in *Disrupting criminal networks*. London: Palgrave Macmillan, pp. 33-58.

Black-Michaud, J. (1980) *Feuding societies*. Oxford: Basil Blackwell.

Blumer, H. (2012) 'Symbolic interactionism', in *Contemporary Sociological Theory*. Hoboken, NJ: John Wiley & Sons, pp. 62-68.

Bock, K., Shannon, S., Movahedi, Y. and Cukier, M. (2017) 'Application of routine activity theory to cyber intrusion location and time', in *2017 13th European Dependable Computing Conference (EDCC)*, pp. 139-146.

Boehm, C. (1987) *Blood revenge: the enactment and management of conflict in Montenegro and other tribal societies*. Philadelphia: University of Pennsylvania Press.

Booth, R. (2024) 'The secret health hell of being scammed: “I felt my mind disintegrating”', *The Guardian*, 23 October. Available at: https://www.theguardian.com/society/2024/oct/23/secret-health-hell-being-scammed-felt-mind-disintegrating (Accessed: 15 August 2025).

Boyko, J. et al. (2023) *An interdisciplinary outlook on large language models for scientific research*. Available at: https://arxiv.org/abs/2311.04929 (Accessed: 15 August 2025).

Boyle, D., DeZoort, F. and Hermanson, D. (2015) 'The effect of alternative fraud model use on auditors’ fraud risk judgments', *Journal of Accounting and Public Policy*, 34(6), pp. 578-596.

Braithwaite, J. (1989) *Crime, shame and reintegration*. Cambridge: Cambridge University Press.

Brenner, S.W. (2007) 'History of computer crime', in *The history of information security*. Amsterdam: Elsevier Science BV, pp. 705-721.

Briggs, A. and Burke, P. (2009) *A social history of the media: from Gutenberg to the Internet*. 3rd edn. Cambridge: Polity Press.

Brown, R.M. (1975) *Strain of violence: historical studies of American violence and vigilantism*. Oxford: Oxford University Press.

Button, M. (2021) 'Hiding behind the veil of Action Fraud…', *Policing: A Journal of Policy and Practice*, 15(2), pp. 1093-1107.

Button, M. (2025) 'Policing cross border fraud “above and below the surface”', *Crime, Law and Social Change*, 83, pp. 1-19.

Button, M. and Cross, C. (2017) *Cyber frauds, scams and their victims*. London: Routledge.

Button, M. and Whittaker, J. (2021) 'Exploring the voluntary response to cyber-fraud: From vigilantism to responsibilisation', *International Journal of Law, Crime and Justice*, 66, 100481.

Callon, M. (2001) 'Actor network theory', in *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier, pp. 62-66.

Campbell, C. and Marks, L.J. (2015) 'Good native advertising isn’t a secret', *Business Horizons*, 58(6), pp. 599-606.

Carter, M. and Fuller, C. (2016) 'Symbols, meaning, and action: the past, present, and future of symbolic interactionism', *Current Sociology*, 64(6), pp. 931-961.

Chang, L.Y. and Zhu, J. (2020) 'Taking justice into their own hands: predictors of vigilantism among cyber citizens in Hong Kong', *Frontiers in Psychology*, 11, p. 556903.

Charmaz, K. (2017) 'The power of constructivist grounded theory for critical inquiry', *Qualitative Inquiry*, 23(1), pp. 34-45.

Charmaz, K. (2020) '"With constructivist grounded theory you can't hide": social justice research and critical inquiry in the public sphere', *Qualitative Inquiry*, 26(2), pp. 165-176.

Checkland, P. (1999) *Systems thinking, systems practice*. Chichester: John Wiley & Sons.

Chen, W., Wang, F. and Edwards, M. (2022) *Active countermeasures for email fraud*. Available at: https://arxiv.org/abs/2210.15043 (Accessed: 15 August 2025).

Chesney, R. and Citron, D. (2019) 'Deepfakes and the new disinformation war: the coming age of post-truth geopolitics', *Foreign Affairs*, 98(1), pp. 147-155.

Chew, R. et al. (2023) *LLM-assisted content analysis: using large language models to support deductive coding*. Available at: https://arxiv.org/abs/2306.14924 (Accessed: 15 August 2025).

Chiou, L. and Tucker, C. (2013) 'Paywalls and the demand for news', *Information Economics and Policy*, 25(2), pp. 61-69.

Choi, K. (2008) 'Computer crime victimization and integrated theory: an empirical assessment', *International Journal of Cyber Criminology*, 2(1), pp. 308-333.

Choo, K.K.R. (2014) 'A conceptual interdisciplinary plug-and-play cyber security framework', in *Cyber Behavior*. New York: Springer, pp. 81-99.

Christie, N. (1986) 'The ideal victim', in *From crime policy to victim policy*. London: Palgrave Macmillan, pp. 17-30.

Cohen, L.E. and Felson, M. (2010) 'Social change and crime rate trends: a routine activity approach (1979)', in *Classics in environmental criminology*. London: Routledge, pp. 203-232.

Costello, B. (2017) 'Social control theory', in *The Springer handbook of social problems*. Cham: Springer, pp. 31-41.

Couldry, N. and Turow, J. (2014) 'Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy', *International Journal of Communication*, 8, pp. 1710-1726.

Craig, D. and Cunningham, S. (2019) *Social media entertainment: the new intersection of Hollywood and Silicon Valley*. New York: NYU Press.

Cressman, D. (2018) 'Actor-network theory', in *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: John Wiley & Sons.

Cross, C. and Blackshaw, D. (2015) 'Improving the Police Response to Online Fraud', *Policing: A Journal of Policy and Practice*, 9(2), pp. 119–128. Available at: https://doi.org/10.1093/police/pau044 (Accessed: 15 August 2025).

Cross, C. and Mayers, D. (2021) 'Scambaiter Narratives of Victims and Offenders and Their Influence on the Policing of Fraud', *Policing: A Journal of Policy and Practice*, 15(4), pp. 2148–2164.

Crown Prosecution Service (2023) *CPS Economic Crime Strategy 2025: Two-Year Progress Report*. Available at: https://www.cps.gov.uk/publication/cps-economic-crime-strategy-2025-two-year-progress-report (Accessed: 15 August 2025).

Cunningham, S. and Craig, D. (2019) 'Creator governance in social media entertainment', *Social Media + Society*, 5(4). Available at: https://doi.org/10.1177/2056305119883428 (Accessed: 15 August 2025).

Cybersecurity Ventures (2023) *Cybercrime To Cost The World $10.5 Trillion Annually By 2025*. Available at: https://cybersecurityventures.com/cybercrime-to-cost-the-world-10-5-trillion-annually-by-2025/ (Accessed: 15 August 2025).

Denning, D.E. (2010) 'Internet is transforming', in *Handbook of Internet Crime*. London: Routledge, p. 194.

DEYOCLUB (2021a) *Scambaiting the scammers: my journey to fighting scammers*. [YouTube video]. 21 April. Available at: https://www.youtube.com/watch?v=F-uh3O5T6aM (Accessed: 15 August 2025).

DEYOCLUB (2021b) *We're back calling scammers!*. [YouTube video]. 16 May. Available at: https://www.youtube.com/watch?v=D-l7Yk-nK4E (Accessed: 15 August 2025).

Doležal, D. (2024) 'Digital vigilantism: examining online justice, ethical challenges and the role of social media', *Hrvatska revija za rehabilitacijska istraživanja*, 60(2), pp. 219–245.

Duffy, B. (2015) 'Amateur, autonomous, and collabourative: myths of aspiring female cultural producers in Web 2.0', *Critical Studies in Media Communication*, 32(1), pp. 48-64.

Dynel, M. (2021) 'On scams, scambaiting, deception, and epistemological vigilance in #scambaiting videos on YouTube', *Social Media + Society*, 7(4). Available at: https://doi.org/10.1177/20563051211055935 (Accessed: 15 August 2025).

Elberse, A. (2013) *Blockbusters: why big hits–and big risks–are the future of the entertainment business*. London: Faber & Faber.

Elder-Vass, D. (2017) 'Actor-network theory', in *The SAGE Encyclopedia of Political Behavior*. Thousand Oaks, CA: SAGE Publications.

Emsley, C. (ed.) (2017) *Theories and origins of the modern police*. London: Routledge.

Europol (2023) *Spotlight Report on Online Fraud Schemes*. Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/Spotlight-Report_Online-fraud-schemes.pdf (Accessed: 15 August 2025).

Evans, D.S. (2009) 'The online advertising industry: economics, evolution, and privacy', *Journal of Economic Perspectives*, 23(3), pp. 37-60.

Favarel-Garrigues, G. (2021) ''Vigilante shows' and law enforcement in Russia', *Europe-Asia Studies*, 73(2), pp. 221-242.

FBI (2025) *2024 Internet Crime Report*. Washington, D.C.: Internet Crime Complaint Center.

Featherstone, R. and Deflem, M. (2003) 'Anomie and strain: context and consequences of Merton's two theories', *Sociological Inquiry*, 73(4), pp. 471-489.

Ferrell, J., Hayward, K. and Young, J. (2008) *Cultural Criminology: An Invitation*. London: Sage.

Finn, B. (2021) 'Perspective: blending public and private law enforcement', *FBI Law Enforcement Bulletin*, 26 May.

Forst, B. (2000) 'The privatisation and civilianization of policing', in *Criminal Justice 2000, Volume 2*. Washington, D.C.: U.S. Department of Justice.

Frampton, L. (2022) ''Paedophile hunters': practitioner perspectives', *Probation Journal*, 70(2), pp. 143-159.

FTC (2023) *FTC crunches 2022 numbers: see where scammers continue to crunch consumers*. Available at: https://www.ftc.gov/business-guidance/blog/2023/02/ftc-crunches-2022-numbers-see-where-scammers-continue-crunch-consumers (Accessed: 15 August 2025).

Gabdulhakov, R. (2018) 'Citizen-led justice in post-communist Russia: from comrades' courts to dotcomrade vigilantism', *Surveillance & Society*, 16(3), pp. 315-329.

Gagarin, M. and Woodruff, P. (2015) 'Early Greek legal thought', in *A Treatise of Legal Philosophy and General Jurisprudence: Volume 6: A History of the Philosophy of Law from the Ancient Greeks to the Scholastics*. Dordrecht: Springer, pp. 7-34.

GASA and Feedzai (2024) *The Global State of Scams 2024*. Available at: https://www.gasa.org/post/global-state-of-scams-report-2024-1-trillion-stolen-in-12-months-gasa-feedzai (Accessed: 15 August 2025)

Gerber, E.M., Hui, J.S. and Kuo, P.Y. (2012) 'Crowdfunding: why people are motivated to post and fund projects on crowdfunding platforms', in *Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics*, vol. 2, no. 11, p. 10.

Gillespie, T. (2010) 'The politics of 'platforms'', *New Media & Society*, 12(3), pp. 347-364.

Goffman, E. (1959) *The presentation of self in everyday life*. New York: Anchor Books.

Goffman, E. (1963) *Stigma: notes on the management of spoiled identity*. Englewood Cliffs, NJ: Prentice-Hall.

Goffman, E. (1967) *Interaction ritual: essays on face-to-face behavior*. New York: Aldine Publishing.

Goffman, E. (1974) *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.

Gomez-Uribe, C.A. and Hunt, N. (2016) 'The Netflix recommender system: algorithms, business value, and innovation', *ACM Transactions on Management Information Systems*, 6(4), pp. 1-19.

Govender, I., Watson, B. and Amra, J. (2021) 'Global virus lockdown and cybercrime rate trends: a routine activity approach', *Journal of Physics: Conference Series*, 1828(1), p. 012107.

Greenhouse, C. (2018) 'Reading Durkheim in darkness', *Law & Humanities eJournal*, 12(2), pp. 193-215.

Groombridge, N. (2018) *Crime and Media: A Reader*. London: Routledge.

Gupta, S., Gupta, P. and Kumaraguru, P. (2015) *Abusing phone numbers and cross-application features for crafting targeted attacks*. Available at: https://arxiv.org/abs/1512.07330 (Accessed: 15 August 2025).

Haas, N.E., de Keijser, J.W. and Bruinsma, G.J. (2014) 'Public support for vigilantism, confidence in police and police responsiveness', *Policing and Society*, 24(2), pp. 224-241. Available at: https://doi.org/10.1080/10439463.2012.725983.

Hadnagy, C. and Fincher, M. (2015) *Phishing dark waters: the offensive and defensive sides of malicious emails*. Indianapolis, IN: John Wiley & Sons.

Harding, A. (1960) 'The origins and early history of the keeper of the peace', *Transactions of the Royal Historical Society*, 10, pp. 85-109.

Hayward, K. (2015) *Five Spaces of Cultural Criminology*. Available at: https://blogs.kent.ac.uk/culturalcriminology/files/2015/03/Five-Spaces-of-Cultural-Criminology.pdf (Accessed: 15 August 2025).

Hayward, K. and Young, J. (2004) 'Cultural criminology: Some notes on the script', *Theoretical Criminology*, 8(3), pp. 259–273.

Henry, S. and Milovanovic, D. (1996) *Constitutive Criminology: Beyond Postmodernism*. London: Sage.

Homer, E. (2019) 'Testing the fraud triangle: a systematic review', *Journal of Financial Crime*, 27(1), pp. 172-187.

Hook, N. (2015) 'Grounded theory', in *The SAGE Encyclopedia of Theory in Criminology and Criminal Justice*. Thousand Oaks, CA: SAGE Publications, pp. 309-320.

Hosseini, M. and Horbach, S. (2023) 'Fighting reviewer fatigue or amplifying bias? considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review', *Research Square*.

Hosseini, M., Resnik, D. and Holmes, K. (2023) 'The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts', *Research Ethics*, 19(4), pp. 449-465.

Hua, Y. et al. (2022) 'Characterizing alternative monetisation strategies on YouTube', *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), pp. 1-30.

Huang, Q. (2021) 'The mediated and mediatised justice-seeking: Chinese digital vigilantism from 2006 to 2018', *Internet Histories*, 5(3-4), pp. 304-322.

Ilan, J. (2019) *Cultural Criminology and the Digital*. London: Routledge.

Jackson, M. (2001) 'Systems thinking and the social sciences', in *Systems thinking for the 21st century*. Boston, MA: Springer, pp. 239-247.

Jane, E.A. (2016) 'Online misogyny and feminist digilantism', *Continuum*, 30(3), pp. 284-297.

Jauregui, B. (2015) 'Just war: the metaphysics of police vigilantism in India', *ARCS*, 1(1), pp. 41-59.

Jenkins, H., Ford, S. and Green, J. (2013) *Spreadable media: creating value and meaning in a networked culture*. New York: New York University Press.

Jensen, R.I.T. (2024) 'Do awareness campaigns reduce financial fraud?', *European Journal on Criminal Policy and Research*, 30, pp. 1-22.

Jim Browning (2021) *Scammer gets a new script*. [YouTube video]. 15 October. Available at: https://www.youtube.com/watch?v=fXpfdqVb2-M (Accessed: 15 August 2025).

Jim Browning (2022a) *Inside a pig butchering scam*. [YouTube video]. 28 October. Available at: https://www.youtube.com/watch?v=p-a2EwB8-tA (Accessed: 15 August 2025).

Jim Browning (2022b) *I got access to a scammer's CCTV cameras*. [YouTube video]. 1 April. Available at: https://www.youtube.com/watch?v=P6dhteJIY48 (Accessed: 15 August 2025).

Johnson, R.M. (2022) 'Social media and free speech: a collision course that threatens democracy', *Ohio Northern University Law Review*, 49(2), p. 461.

Johnston, L. (1996) 'What is vigilantism?', *British Journal of Criminology*, 36(2), pp. 220-236.

Johnston, L. (2005) *The rebirth of private policing*. London: Routledge.

Jordan, T. and Taylor, P. (2004) *Hacktivism and cyberwars: rebels with a cause?*. London: Routledge.

Kaafar, D. et al. (2024) 'Real criminals, fake victims: how chatbots are being deployed in the global fight against phone scammers', *The Guardian*, 6 July.

Karjus, A. (2023) *Machine-assisted mixed methods: augmenting humanities and social sciences with artificial intelligence*. Available at: https://arxiv.org/abs/2309.14379 (Accessed: 15 August 2025).

Keating, C., Katina, P., Hodge, R. and Bradley, J. (2020) 'Systems theory: bridging the gap between science and practice for systems engineering', in *INCOSE International Symposium*, 30(1), pp. 1113-1128.

Kigerl, A. (2012) 'Routine activity theory and the determinants of high cybercrime countries', *Social Science Computer Review*, 30(4), pp. 470-486.

Kitboga (2017) *The angriest scammer I've ever called (do not redeem)*. [YouTube video]. 13 November. Available at: https://www.youtube.com/watch?v=SNz4m8p82ks (Accessed: 15 August 2025).

Kitboga (2020) *I called scammers as two different people*. [YouTube video]. 13 January. Available at: https://www.youtube.com/watch?v=3om0-fW4384 (Accessed: 15 August 2025).

Kitboga (2021a) *I trapped 200 scammers in an impossible maze*. [YouTube video]. 19 July. Available at: https://www.youtube.com/watch?v=8_3h-12nRsA (Accessed: 15 August 2025).

Kitboga (2021b) *Raging scammers waste 54 hours on me (Crow Pro 1)*. [YouTube video]. 23 August. Available at: https://www.youtube.com/watch?v=2VhGIvD-W-E (Accessed: 15 August 2025).

Kitboga (2021c) *I made scammers think I'm a big deal*. [YouTube video]. 20 September. Available at: https://www.youtube.com/watch?v=0sa2-vjS_Is (Accessed: 15 August 2025).

Kitboga (2021d) *I pretended to be a scammer to a scammer*. [YouTube video]. 1 November. Available at: https://www.youtube.com/watch?v=aG0eS2f0cv8 (Accessed: 15 August 2025).

Kitboga (2021e) *The most evil scammer I've ever called*. [YouTube video]. 20 December. Available at: https://www.youtube.com/watch?v=p4843z8_2s8 (Accessed: 15 August 2025).

Kitboga (2022a) *When scammers see $0.00 in your bank....* [YouTube video]. 11 April. Available at: https://www.youtube.com/watch?v=W-Lvi-PaBwY (Accessed: 15 August 2025).

Kitboga (2022b) *Scammers think I'm a grandma who loves bingo*. [YouTube video]. 14 February. Available at: https://www.youtube.com/watch?v=u8-3f9y-h5w (Accessed: 15 August 2025).

Kitboga (2022c) *Spending all my money while scammers watch*. [YouTube video]. 20 June. Available at: https://www.youtube.com/watch?v=yYAjL0C3t-M (Accessed: 15 August 2025).

Kitboga (2022d) *Scammer pretends to be my dead wife*. [YouTube video]. 25 July. Available at: https://www.youtube.com/watch?v=JV9gPNQghN0 (Accessed: 15 August 2025).

Kumar, S. and Sethi, S. (2007) 'Dynamic pricing and advertising for web content providers', *European Journal of Operational Research*, 197(3), pp. 924-944.

Lazarus, S., Whittaker, J., McGuire, M. and Platt, L. (2023) 'What do we know about online romance fraud studies? A systematic review of the empirical literature (2000 to 2021)', *Journal of Economic Criminology*, 2, 100013.

Lea, J. (2015) 'Jock Young and the development of left realist criminology', *Critical Criminology*, 23(2), pp. 165-177.

Legocki, K., Walker, K. and Kiesler, T. (2020) 'Sound and fury: digital vigilantism as a form of consumer voice', *Journal of Public Policy & Marketing*, 39(2), pp. 169-187.

Leukfeldt, E.R. (2016) 'Applying routine activity theory to cybercrime', *Journal of Research in Crime and Delinquency*, 53(4), pp. 560–581.

Li, Z. (2023) *The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination*. Available at: https://arxiv.org/abs/2304.14347 (Accessed: 15 August 2025).

Lin, Z. (2023) 'Why and how to embrace AI such as ChatGPT in your academic life', *Royal Society Open Science*, 10(7), p. 230658.

Liu, Y. et al. (2023) *Summary of ChatGPT/GPT-4 research and perspective towards the future of large language models*. Available at: https://arxiv.org/abs/2304.01852 (Accessed: 15 August 2025).

Loader, I. (2013) 'Why do the police matter?: beyond the myth of crime-fighting', in *The future of policing*. London: Routledge, pp. 40-51.

Loader, I. and Walker, N. (2007) *Civilising security*. Cambridge: Cambridge University Press.

Lobato, R. (2016) 'The cultural logic of digital intermediaries: YouTube multichannel networks', *Convergence*, 22(4), pp. 348-360.

Lokanan, M. (2015) 'Challenges to the fraud triangle: questions on its usefulness', *Accounting Forum*, 39(3), pp. 201-224.

Losty, P.A. (1976) *An introduction to general systems thinking*. London: Open University Press.

Loveluck, B. (2016) 'Digital vigilantism, between denunciation and punitive action', *Politix*, 114(2), pp. 127-153.

Loveluck, B. (2019) 'The many shades of digital vigilantism: a typology of online self-justice', *Global Crime*, 21(3-4), pp. 213-241.

Lyng, S. (1990) 'Edgework: A social psychological analysis of voluntary risk taking', *American Journal of Sociology*, 95(4), pp. 851–886.

Makridakis, S., Petropoulos, F. and Kang, Y. (2023) 'Large language models: their success and impact', *Forecasting*, 5(3), pp. 419-431.

Malcolm Merlyn (2018) *Scammer vs Anonymous (fake CMD, event viewer, syskey...)*. [YouTube video]. 13 May. Available at: https://www.youtube.com/watch?v=qTz3iS52h4g (Accessed: 15 August 2025).

Malcolm Merlyn (2021) *Scammer call center gets destroyed by call flooding*. [YouTube video]. 29 January. Available at: https://www.youtube.com/watch?v=y2F0-P8t9nI (Accessed: 15 August 2025).

Mann, S. et al. (2023) 'AUTOGEN: a personalized large language model for academic enhancement, ethics and proof of principle', *The American Journal of Bioethics*, 23(9), pp. 28-41.

Martínez-López, F.J., Li, Y. and Young, S.M. (2022) *Social media monetisation: platforms, strategic models and critical success factors*. Cham: Springer Nature.

McBain, G. (2015) 'Modernising the law: breaches of the peace & justices of the peace', *Journal of Policing & Law*, 8(1), p. 158.

McGuire, M. (2012) *Technology, Crime, and Justice*. London: Routledge.

McGuire, M.R. (2016) 'Cybercrime 4.0: now what is to be done?', in Jeffery, C. (ed.) *What is to be done about crime and punishment? towards a 'public criminology'*. London: Palgrave Macmillan, pp. 251-279.

McGuire, M.R. (2021) 'The laughing policebot: automation and the end of policing', *Policing and Society*, 31(1), pp. 20-36.

McKay, T. (2022) *Scambaiting is racist and dangerous – stop celebrating it*. The Next Web, 15 March. Available at: https://thenextweb.com/news/scambaiting-racist-dangerous-stop-celebrating-syndication (Accessed: 15 August 2025).

McKenna, N. et al. (2023) 'Sources of hallucination by large language models on inference tasks', in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 2758-2774.

Meadows, D.H. (2008) *Thinking in systems: a primer*. White River Junction, VT: Chelsea Green Publishing.

Mele, C., Pels, J. and Polese, F. (2010) 'A brief review of systems theories and their managerial applications', *Service Science*, 2(1-2), pp. 126-135.

Metropolitan Police (2022) *iSpoof fraud investigation*. Available at: https://news.met.police.uk/news/ispoof-fraud-investigation-457700 (Accessed: 15 August 2025).

Milligan, I. (2023) 'The modem world: a prehistory of social media by Kevin Driscoll', *Technology and Culture*, 64(3), pp. 1000-1001.

Mills, J., Bonner, A. and Francis, K. (2006a) 'Adopting a constructivist approach to grounded theory: implications for research design', *International Journal of Nursing Practice*, 12(1), pp. 8-13.

Mills, J., Bonner, A. and Francis, K. (2006b) 'The development of constructivist grounded theory', *International Journal of Qualitative Methods*, 5(1), pp. 25-35.

Mollick, E. (2014) 'The dynamics of crowdfunding: an exploratory study', *Journal of Business Venturing*, 29(1), pp. 1-16.

Moloney, C.J., Unnithan, P., & Zhang, W. (2022) 'Assessing law enforcement's cybercrime capacity and capability', *FBI Law Enforcement Bulletin*, 10 October.

Montañez, A. et al. (2020) 'Cognitive biases and susceptibility to phishing', *Computers in Human Behavior*, 109, 106349.

Montgomery, R. and Griffiths, C.T. (2016) *The use of private security services for policing. Research Report 2015–R041*. Ottawa: Public Safety Canada.

More Kitboga (2021a) *I made an entire call center angry over losing $94,000*. [YouTube video]. 22 September. Available at: https://www.youtube.com/watch?v=N4N22m8j1g0 (Accessed: 15 August 2025).

More Kitboga (2021b) *Angriest scammer Kitboga has ever called (full 10 hours)*. [YouTube video]. 15 December. Available at: https://www.youtube.com/watch?v=b2-yex_3g3I (Accessed: 15 August 2025).

More Kitboga (2022) *The funniest scammer I've ever called (full 10 hours)*. [YouTube video]. 2 May. Available at: https://www.youtube.com/watch?v=0fICIJo3hjc (Accessed: 15 August 2025).

Mui, G. and Mailley, J. (2015) 'A tale of two triangles: comparing the fraud triangle with criminology's crime triangle', *Accounting Research Journal*, 28(1), pp. 45-58.

Muhammad, M. and Mukhtar, J. (2015) 'Social control: genesis, conceptual, and theoretical issues', *International Journal of Innovative Research and Development*, 4(8), pp. 132-137.

National Trading Standards (2023) *19 million lose money to scams but fewer than a third report*. Available at: https://www.nationaltradingstandards.uk/news/19-million-lose-money-to-scams-but-fewer-than-a-third-report/ (Accessed: 15 August 2025).

Nissenbaum, H. (2011) 'A contextual approach to privacy online', *Daedalus*, 140(4), pp. 32-48.

NordVPN (2023) *What is scambaiting?*. Available at: https://nordvpn.com/blog/scambaiting/ (Accessed: 15 August 2025).

OECD (2023) *Consumer vulnerability in the digital age. OECD Digital Economy Papers, No. 355*. Paris: OECD Publishing.

Ofcom (2023) *Scale and impact of online fraud revealed*. Available at: https://www.ofcom.org.uk/news-centre/2023/scale-and-impact-of-online-fraud-revealed (Accessed: 15 August 2025).

Orcutt, J. (2016) 'Crime, social control theory of', in *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: John Wiley & Sons, pp. 1-4.

Ormrod, W.M., Dodd, G. and Musson, A. (eds.) (2009) *Medieval petitions: grace and grievance*. Woodbridge: Boydell & Brewer.

Parsons, E.F. (2015) *Ku-Klux: the birth of the Klan during Reconstruction*. Chapel Hill, NC: UNC Press Books.

Perkins, M. (2023) 'Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond', *Journal of University Teaching and Learning Practice*, 20(2), p. 07.

Pleasant Green (2022) *I tricked a scammer into hacking me. here's what I found...* [YouTube video]. 26 May. Available at: https://www.youtube.com/watch?v=zQzaMv5usC4 (Accessed: 15 August 2025).

Pratt, T.C., Holtfreter, K. and Reisig, M.D. (2010) 'Routine online activity and internet fraud targeting: extending the generality of routine activity theory', *Journal of Research in Crime and Delinquency*, 47(3), pp. 267-296.

Presdee, M. (2000) *Cultural Criminology and the Carnival of Crime*. London: Routledge.

Prothero, I. (2013) *Artisans and politics in early nineteenth-century London (Routledge Revivals): John Gast and his times*. London: Routledge.

Rabie, M. (2013) 'Processes of societal transformation', in *Global economic and political governance*. London: Palgrave Macmillan, pp. 43-57.

Rawlings, P. (2012) 'Policing before the police', in *Handbook of policing*. London: Routledge, pp. 75-99.

Reuters (2025) 'FBI says cybercrime costs rose to at least $16 billion in 2024', *Reuters*, 23 April.

Rieder, B. (2023) 'Making a living in the creator economy: a large-scale analysis of monetisation and networking strategies', *Social Media + Society*, 9(3). Available at: https://doi.org/10.1177/20563051231196658.

Rinoa Antidote (2021) *Scammer gets angry over fake bank*. [YouTube video]. 25 October. Available at: https://www.youtube.com/watch?v=p0Wz3xA3bqs (Accessed: 15 August 2025).

Rinoa Poison (2021) *Irritating scammers!*. [YouTube video]. 25 January. Available at: https://www.youtube.com/watch?v=fWz-pMMGx-I (Accessed: 15 August 2025).

Romanosky, S. (2016) 'Examining the costs and causes of cyber incidents', *Journal of Cybersecurity*, 2(2), pp. 121-135.

Ross, A.S. and Logi, L. (2021) ''Hello, this is Martha': Interaction dynamics of live scambaiting on Twitch', *Convergence*, 27(6), pp. 1789-1810. Available at: https://doi.org/10.1177/13548565211015453.

Rußell, R. et al. (2020) 'Monetising online content: digital paywall design and configuration', *Business & Information Systems Engineering*, 62(3), pp. 253-260.

Sauter, M. (2014) *The coming swarm: DDOS actions, hacktivism, and civil disobedience on the Internet*. New York: Bloomsbury Academic.

SCARS (2021) *SCARS position statement against scambaiting*. Romance Scams Now. Available at: https://romancescamsnow.com/dating-scams/scars-position-statement-against-scambaiting/ (Accessed: 15 August 2025).

Scammer Payback (2021a) *17 minutes of this **INSANE** scammer raging*. [YouTube video]. 15 October. Available at: https://www.youtube.com/watch?v=t-Rv9y88a-E (Accessed: 15 August 2025).

Scammer Payback (2021b) *Moments before scammers win*. [YouTube video]. 22 April. Available at: https://www.youtube.com/watch?v=Cq_7h2wXwII (Accessed: 15 August 2025).

Scammer Payback (2021c) *Telling scammers their address*. [YouTube video]. 29 April. Available at: https://www.youtube.com/watch?v=vGAbZXB2OaM (Accessed: 15 August 2025).

Scammer Payback (2021d) *Scammer's reaction after being hacked*. [YouTube video]. 29 October. Available at: https://www.youtube.com/watch?v=u1a433-d0fM (Accessed: 15 August 2025).

Scammer Payback (2022) *$4,000 saved! scammer file deletion and PC access*. [YouTube video]. 22 January. Available at: https://www.youtube.com/watch?v=ZfQ4-A1G44M (Accessed: 15 August 2025).

ScammerRevolts (2018) *Deleting every file off a scammers laptop!*. [YouTube video]. 10 August. Available at: https://www.youtube.com/watch?v=z8zDqQxM9aA (Accessed: 15 August 2025).

Schreck, C. (2014) 'Social control theories', in *The Encyclopedia of Criminology and Criminal Justice*. Hoboken, NJ: John Wiley & Sons, pp. 1-8.

Schuchter, A. and Levi, M. (2015) 'Beyond the fraud triangle: Swiss and Austrian elite fraudsters', *Accounting Forum*, 39(3), pp. 176-187.

Schuchter, A. and Levi, M. (2016) 'The fraud triangle revisited', *Security Journal*, 29(2), pp. 107-121.

Silke, A. (2001) 'Dealing with vigilantism: issues and lessons for the police', *The Police Journal: Theory, Practice and Principles*, 74(2), pp. 120-133.

Silva, K. (2018) 'Vigilantism and cooperative criminal justice: is there a place for cybersecurity vigilantes in cybercrime fighting?', *International Review of Law, Computers & Technology*, 32(1), pp. 1-16.

Simon, H.A. (1962) 'The architecture of complexity', *Proceedings of the American Philosophical Society*, 106(6), pp. 467-482.

Smallridge, J., Wagner, P. and Crowl, J.N. (2016) 'Understanding cyber-vigilantism: a conceptual framework', *Journal of Theoretical & Philosophical Criminology*, 8(1), pp. 58-81.

Smith, K.S. (2014) *Emile Durkheim and the collective consciousness of society: a study in criminology*. London: Anthem Press.

Smith, O. and Raymen, T. (2016) 'Deviant leisure: A criminological perspective', *Theoretical Criminology*, 20(3), pp. 283–299.

Solomon, F. (2024) '"Pig butchering" scams cost Americans billions. this lawyer is taking them on', *The Wall Street Journal*, 17 September.

Song, J., Kim, H. and Gkelias, A. (2014) 'iVisher: real-time detection of caller ID spoofing', *ETRI Journal*, 36(5), pp. 865-875.

Sorell, T. (2019) 'Scambaiting on the spectrum of digilantism', *Criminal Justice Ethics*, 38(3), pp. 153-175.

Sorunke, O. (2016) 'Personal ethics and fraudster motivation: the missing link in fraud triangle and fraud diamond theories', *The International Journal of Academic Research in Business and Social Sciences*, 6(2), pp. 159-165.

Soulliére, D. (1999) *Police and technology: historical review and current status*. Ottawa: Canadian Police College.

Starn, O. (1999) *Nightwatch: the politics of protest in the Andes*. Durham, NC: Duke University Press.

Steinmetz, K.F. and Holt, T.J. (2022) 'Falling for social engineering: a qualitative analysis of social engineering policy recommendations', *Social Science Computer Review*, 40(5), pp. 1198-1216.

Tai, R. et al. (2023) *Use of large language models to aid analysis of textual data*. bioRxiv.

Tan, X.W., See, K. and Kok, S. (2024) *ScamGPT J: inside the scammer's mind, a generative AI-based approach toward combating messaging scams*. Available at: https://arxiv.org/abs/2402.13528 (Accessed: 15 August 2025).

Tanner, S. and Campana, A. (2020) '"Watchful citizens" and digital vigilantism: a case study of the far right in Quebec', *Global Crime*, 21(3-4), pp. 262-282.

Tarr, J.A. (1992) 'The city and the telegraph: urban telecommunications in the pre-telephone era', *Journal of Urban History*, 14(1), pp. 38-80.

Täuscher, K. and Laudien, S.M. (2018) 'Understanding platform business models: a mixed methods study of marketplaces', *European Management Journal*, 36(3), pp. 319-329.

Territoriality (2025) *Investigating cybercrime: the key jurisdictional and technical challenges faced by law enforcement and ways to address them*. York: University of York.

***The Guardian* (2021)** 'Who scams the scammers? meet the amateur scambaiters taking on the crooks', *The Guardian*, 3 October. Available at: https://www.theguardian.com/technology/2021/oct/03/who-scams-the-scammers-meet-the-amateur-scambaiters-taking-on-the-crooks (Accessed: 15 August 2025).

The Hoax Hotel (2021) *Scammers don't want you to know this*. [YouTube video]. 20 October. Available at: https://www.youtube.com/watch?v=sUoDuMKq99g (Accessed: 15 August 2025).

Thompson, E.P. (2016) *The making of the English working class*. New York: Open Road Media.

Trilogy Media (2019) *Confronting a scammer (he cried)*. [YouTube video]. 19 November. Available at: https://www.youtube.com/watch?v=tz-e8oz3hjc (Accessed: 15 August 2025).

Trilogy Media (2021) *Saving victims from scammers in real time*. [YouTube video]. 26 February. Available at: https://www.youtube.com/watch?v=i2K4_nAh2I8 (Accessed: 15 August 2025).

Trilogy Vault (2021) *Scammer money mule thinks she's going on a date with Art*. [YouTube video]. 12 August. Available at: https://www.youtube.com/watch?v=nm8jURtxX_s (Accessed: 15 August 2025).

Trottier, D. (2017) 'Digital vigilantism as weaponisation of visibility', *Philosophy & Technology*, 30(1), pp. 55-72.

Ugwudike, P. (2024) 'Algorithmic injustice: an intersectional perspective on AI in criminal justice', *Criminology & Criminal Justice*, 24(1), pp. 22-39.

UK Finance (2025) *Take five to stop fraud: team talk tactics to help you stay safe from fraud*. [Press release]. 12 May.

UNODC (2025) *Inflection point: global implications of scam centres in Southeast Asia*. Vienna: United Nations Office on Drugs and Crime.

Value4AI (2023) *Awesome LLM in social science*. GitHub. Available at: https://github.com/Value4AI/Awesome-LLM-in-Social-Science (Accessed: 15 August 2025).

Varshney, G., Misra, M. and Atrey, P.K. (2016) 'A survey and classification of web phishing detection schemes', *Security and Communication Networks*, 9(18), pp. 6259-6284.

Wagner, T.M., Benlian, A. and Hess, T. (2014) 'Converting freemium customers from free to premium, the role of the perceived premium fit in the case of music as a service', *Electronic Markets*, 24(4), pp. 259-268.

Wall, D.S. (2024) *Cybercrime: The Transformation of Crime in the Information Age*. 2nd edn. Cambridge: Polity.

Weber, M. (2014) 'Rational-legal authority and bureaucracy', in *Policy Process*. London: Routledge, pp. 323-327.

Whitty, M.T. (2015) 'Anatomy of the online dating romance scam', *Security Journal*, 28(4), pp. 443-455.

Whitty, M.T. (2019) 'Predicting susceptibility to cyberfraud victimhood', *Journal of Financial Crime*, 26(1), pp. 277-292.

Williams, S. (2007) 'Potential spaces of crime: The playful, the destructive and the distinctively human', *Social & Legal Studies*, 16(2), pp. 271–287.

Wood, M.A., et al. (2023) 'Impersonating authority: An analysis of social engineering tactics in cybercrime', *Crime & Delinquency*, 69(4), pp. 745-769.

World Economic Forum (WEF) (2024) *Global Risks Report 2024*. Geneva: World Economic Forum.

Xiao, Z. et al. (2023) 'Supporting qualitative analysis with large language models: combining codebook with GPT-3 for deductive coding', in *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 110-113.

Ye, W. et al. (2023) *Assessing hidden risks of LLMs: an empirical study on robustness, consistency, and credibility*. Available at: https://arxiv.org/abs/2305.10235 (Accessed: 15 August 2025).

Zingerle, A. (2014) 'Towards a categorisation of scambaiting strategies against online advance fee fraud', *International Journal of Art, Culture, Design and Technology (IJACDT)*, 4(2), pp. 39-50.

Zizumbo-Colunga, D. (2017) 'Community, authorities, and support for vigilantism: experimental evidence', *Political Behavior*, 39(4), pp. 989-1015.

Zuboff, S. (2019) 'Surveillance capitalism and the challenge of collective action', *New Labour Forum*, 28(1), pp. 10-29.