Design of Virtual Worlds

Zhenzhen Qi[1]

As we transition from Web 2.0 to Web 3.0, digital interactive narratives are rapidly entering spaces of interactive and experiential design. An increasing number of general audiences are experiencing works of art through interactive installation, internet art, virtual reality (VR), and augmented reality-(AR) based software applications. At the same time, an increasing number of academic curriculums are being redesigned into networked narrative environments, where students are expressing themselves and being entertained while learning new knowledge at the same time. Meanwhile, designers and media theorists are also entering heated debates about what it means to make a story digitally interactive. Central to this debate is a creator-player-computation entanglement that emerges from the audience being able to enact the part of the original narrative through real-time software and hardware interfaces. When a narrative initially shaped by the author is rearranged by the audience through a series of button clicks and subsequently subject to system-level rules automated by computer algorithms, whose story does it become? In this chapter, the researcher investigates the conflict of agency in digital interactive media through historical events and contemporary case studies and inquires about conditions for collective authenticity against the backdrop of computational mediation.

— *Dr. Zhenzhen Qi, US/China*

[1] University of Connecticut, US/China.

The Paradox of Immersion

As video game graphical technologies improve yearly, video games grow closer to immersing players in highly realistic virtual worlds. Starting from the 1970s, video games as an industry have evolved from the primordial elements of Pong into the culture-defining mediums demonstrated by early cinematic games such as Call of Duty and Grand Theft Auto. From then on, cinematic games continued to achieve staggering commercial success, accumulating billions of dollars in the following decades. In online video game forums like Reddit and Discord, players frequently share excitement about iconic big-budget cinematic games such as Red Dead Redemption, Metal Gear Solid, and Tomb Raider. In particular, the Uncharted series is one of the most commercially successful 3D adventure game franchises ever. The game features dynamic, lifelike characters in believable three-dimensional worlds that rival Hollywood action-adventure films and intricate gameplay mechanics closely modeled after real-life outdoor recreational experiences such as rock climbing, speed racing, and more. Several online game communities have widely credited the series for significantly raising the standards of single-player games.[2]

However, in online Reddit forums where players exchange experiences of playing the game, opinions are divided. On one hand, some players consider cinematic qualities to be the main currency in the virtual world. In-game characters with high-definition freckles make it difficult for players to move their eyes away. They feel convinced to invest their time and attention in the virtual world because they instinctively feel that Nathan Drake, the main player character, is alive. They believe they stand beside Nathan, breathing the same air and marveling at the same mountains below their feet. However, other seasoned players increasingly question using powerful computational engines to simulate an alternative world that looks and functions like the one we already inhabit. They feel like they are inside an immersive virtual world that evolves with mundanely simple clicks. If the goal is for players to feel realistic, why ask them to click buttons and remind them that they are players who sit outside the screen with a plastic joystick and controller in hand?

Alex Galloway, an American media theorist, argues that technology is social before it is technical. The user interface is not simply a neutral tool that

[2] Chris Plante, "Uncharted 4 Is the Best (and Possibly Last) Game of Its Kind," The Verge, May 10, 2016, www.theverge.com/2016/5/10/11639246/uncharted-4-cinematic-game-review-ps4-playstation.

https://doi.org/10.14361/9783839471425-010 https://www.inlibra.com/de/agb - Open Access -

facilitates communication between humans and machines but a fundamental aspect of how power and control operate in contemporary society. He coined the term "interface effect"[3] to describe how interfaces shape and mediate our interactions with technology and how they shape our understanding of the virtual world. According to Galloway, interfaces are not just technical objects but also cultural and political ones deeply embedded in social relationships and power structures. He argues that interfaces operate on multiple levels, including as physical devices, software, and user experience. Social media platforms that keep users engaged and generate revenue through engaging and captivating users with pervasive technologies.

Digital interfaces are not neutral or transparent but embedded in larger social, cultural, and political contexts. Designers of virtual experiences must be aware of these contexts to understand the impact of interfaces on our lives and to work towards creating interfaces that promote equity, justice, and democracy.

Designing For Procedurality

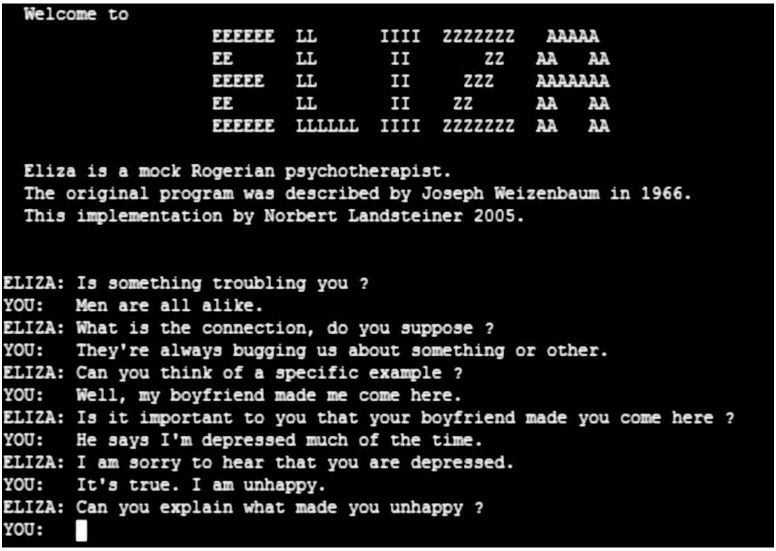
Like Galloway, several other game scholars and practitioners have argued that the real power of videogame design lies above realness or immersion. American game designer Ian Bogost thinks that audiences of print literature and cinema cannot fail a book or movie the way a player fails a videogame. Experiences of games embody experiences of failure.[4]Similarly, Danish game designer Jesper Juul claims, "A video game is half-real: we play by real rules while imagining a fictional world. We win or lose the game in the real world, but we slay a dragon (for example) only in the world of the game."[5] He believes that videogame can simultaneously embody two modes of expression—telling a story and interacting with a set of procedural rules. Therefore, playing a video game immerses one in a fictional world while embracing the natural consequences of its actions just as in the real world. Espen J. Aarseth, a Norwegian scholar specializing in the game study and electronic literature,

[3] Alexander R. Galloway, *The Interface Effect.* (Malden: Polity Press, 2012).vii

[4] Ian Bogost, *How to Do Things with Videogames*, (University of Minnesota Press, 2011), 126–128.

[5] Jesper Juul, *Half-Real: Video Games between Real Rules and Fictional Worlds*, (Cambridge: The MIT Press, 2011).

also referred to hypertext literature, a form of nonlinear storytelling that requires significant audience effort to actuate the narrative outcome, as an example of a text that is impossible to be read but must be actively "played."[6]



```
  Welcome to
                  EEEEEE  LL      IIII  ZZZZZZZ  AAAAA
                  EE      LL       II       ZZ  AA   AA
                  EEEEE   LL       II     ZZZ   AAAAAAA
                  EE      LL       II    ZZ     AA   AA
                  EEEEEE  LLLLLL  IIII  ZZZZZZZ AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

A conversation with the ELIZA chatbot.[7]

For example, ELIZA was one of the first chatterbots developed against the historical backdrop of personal computing. It is one of the earliest natural language processing programs created by Joseph Weizenbaum at MIT's Artificial Intelligence Laboratory in the mid-1960s. It also served as an early test case for the Turing Test, a test of a machine's ability to exhibit intelligent behavior equivalent to or indistinguishable from humans. If one asks a few complex questions, ELIZA fails very quickly by today's standards. However, people found it attractive when it was ported to a PC. This program

[6] Espen J. Aarseth, "Nonlinearity and Literary Theory," 768–770. In *The New Media Reader*, ed. Noah Wardrip-Fruin and Nick Montfort (Cambridge, MA: The MIT Press, 2003), 762–80.

[7] *A Conversation with the ELIZA Chatbot*, accessed June 15, 2022, https://commons.wikimedia.org/wiki/File:ELIZA_conversation.png.

https://doi.org/10.14361/9783839471425-010 https://www.inlibra.com/de/agb - Open Access -

uses "pattern matching" and replacement methods to provide a stereotyped response that makes it feel like an early user is talking to someone who understands their input. ELIZA's most famous iteration was called DOCTOR. It responds like a Rogerian psychotherapist, who "reflects" the question by returning it to the patient.

ELIZA is one example that demonstrates the difference between linearity and nonlinearity in digital interactive experiences mediated by a computer algorithm. Aarseth clarified the difference between linear and nonlinear mediums. More specifically, he defines nonlinearity through a list of variables: Topology, Dynamics, Determinability, Transiency, Maneuverability, and User-functionality. Topology states that nonlinearity does not present an experience from a singular, stable, or sequential vantage point. Instead, through player-text reciprocity enabled by cybernetic agency, a random sequence of the text emerges with each new appointment. Dynamics refers to if the content of the text is constant (reading a print newspaper) or changes (like chatting with ELIZA). Determinability refers to relations between the adjacent content. For example, in natural language processing, there is a higher probability of certain words appearing next to each other, such as "I" and "am," but it is not absolute or deterministic. Likewise, the conversation outcome from chatting with ELIZA is also unpredictable or deterministic. Transiency refers to whether the passage of time alone causes the digital experience to shift. Maneuverability concerns how easily users can access the technical infrastructure that sustains the visual and interactive effects. Finally, user functionality refers to the user's ability to provide real-time input and alter the digital experience by exploring the fictional world space, engaging in virtual role-playing by selecting an avatar's action in the world, or engaging with the world poetically by appreciating the virtual scenes or introspecting on one's existence within the virtual world. Enabled by the nonlinearity framework described above, ELIZA offers a way of making sense of the world by reflecting on one's experiences from a distance.

According to philosopher Nelson Goodman, designing a virtual world is not about faithfully recreating the experience of the analog world in the simulated world but "building a world from others,"[8] similar to building a construction map. A map does not grasp what is already there. A good map systematizes a field and contributes to its disclosure. It delivers a

[8] Nelson Goodman, *Ways of Worldmaking*, Hackett Classic 51 (Indianapolis: Hackett, 2013), 7.

perspective that allows players to make meaning from experiences through world disclosure and world orientation—it aims not at understanding the world for what it is, but at understanding our ways of making sense of the world. Goodman suggested that the design of a world is not a fixed or objective reality but is constantly created and recreated through symbols and representations. He argued that we use various symbols, such as language, art, and science, to develop and interpret the world around us. These systems of symbols are not simply mirrors of reality but actively involve creating and shaping our understanding of the world. Inhabitants of a world should think of themselves as ”world makers” who constantly design and recreate the world through symbols and representations. The best way to understand the nature of reality is to recreate one and interpret it against other models of reality.

Systems Design

In the narratology tradition, the hero's journey or the monomyth is a term that describes the tale of a hero. She leaves the comfort of her homestead, with her faith and fortitude trailed through temptations, crises, and catastrophes inside the metaphorical belly of the whale. Finally, she comes home, changed, and ready to change the world around her. In *The Hero with a Thousand Faces*, a book about comparative mythology that also draws from Carl Jung's analytical theory, American writer Joseph Campbell describes this narrative template as, "A hero ventures forth from the world of common day into a region of supernatural wonder: fabulous forces are there encountered, and a decisive victory is won: The hero comes back from this mysterious adventure with the power to bestow boons on his fellow man."[9] In a video game, the hero's halo is transferred to the player through digital interfaces such as mice, buttons, and joysticks. Ask any gamer and they can recount how they spent hours perfecting their look in a video game with their virtual avatar but sometimes ironically played the actual game for only a few minutes. Avatars can be the key to offering more intense and satisfying game experiences. They can increase the feeling of being transported to another world, provide an enhanced sense of agency, and satisfy the need to feel connected to other players and non-player characters. Ten years after releasing the iconic adventure game *Journey*, partly inspired

[9] Joseph Campbell, *The Hero with a Thousand Faces*, 3rd ed. (Novato: New World Library, 2008), 30.

by Joseph Campbell's writing, the game developer recounted the unexpected effect of role-playing via virtual characters, "When they played through the game together, it helped them to grieve. It helped them to let it go, knowing their loved one was going to a better place. I never thought the game would have the power to be essentially therapeutic, to help people, but it has changed many people's lives, and that is the biggest surprise to me."[10]

However, studies have shown that agency extends beyond real to virtual avatars. An effect in the reverse direction also exists, whereby players unconsciously conform to their avatar's expectations and other environmental elements' appearances. Avatars shape their owners. This phenomenon is termed the "Proteus Effect," after the Greek god who could change his physical form.[11] Avatars are not just ornaments—they subject players to a virtual narrative governed by procedural rules that shape the identity of real players embodied in that world. In other words, the proteus effect describes the phenomenon where people will change their in-game behavior based on how they think others expect them to behave—the conformity to norms in a social system.

A system is not static. It is a dynamic entity comprising interdependent parts, members, or agents. The word system first appeared in publications in 1948. Biologist Ludwig Von Bertalanffy used the term to describe the various organismic scientific phenomena he observed. The human body is one of the most ubiquitous biological systems we encounter daily. When we feel cold, our muscles shiver to generate heat and warm our bodies. When we are hot, we sweat and evaporate heat to cool us. Without paying attention, our body automatically maintains a standard temperature range to comfort us. Since this type of system always takes action to cancel out excessive effects and return the current state to its norm, we refer to the canceling process as negative feedback. Systems involving negative feedback tend to resist change and maintain a stable internal environment. System theory refers to this tendency as homeostasis, and the stabilizing state is equilibrium. Besides natural science, negative feedback is widely adopted in engineering processes and machines. For example, there is a cruise system built into cars. It uses

[10] Brendan Sinclair, "Ten Years Later, Jenova Chen Reflects on Journey," *GamesIndustry*, May 10, 2022, https://www.gamesindustry.biz/articles/2022-03-10-ten-years-later-jenova-chen-reflects-on-journey.

[11] Jim Blascovich and Jeremy Bailenson, *Infinite Reality: The Hidden Blueprint of Our Virtual Lives* (New York: William Morrow Paperbacks, 2012).163-172.

control actions to ensure a stable driving speed without delay or overshoot. Cybernetics is the science of exploring regulatory, purposive, and normalizing systems.

In 1975, French historian and philosopher Michel Foucault published *Discipline and Punish*, a genealogical study on imprisonment, a subtle but effective way of social normalization. According to Foucault, observation is an integral method of confinement. While more explicit forms of conformity, such as physical torture, respond to specific actions, observation responds to the lack of actions.[12] Observation as a control mechanism is demonstrated by the Panopticon, a type of architecture for modern imprisonment designed by English philosopher and social theorist Jeremy Bentham in the eighteenth century. Derived from the Greek word for "all-seeing," a panopticon is a multilayered, cylindrical-shaped building. Individual cells occupy the outermost layer of the building, separated by concrete walls. An inner layer of observation corridors allows the correctional officers to patrol each cell and ascend or descend to different floors. Blinds separate the outer and inner layers, allowing the observers to be concealed from the observed. The Panopticon design is conceptually significant because it enabled a kind of invisible gaze from nowhere—an artificially constructed, godlike omnipresence. Under this gaze, inmates do not know precisely when they are being observed, so they act as if they are being watched all the time, even when not. The result is seemingly self-motivated and self-regulated behaviors—the illusion of agency.

Contrary to negative feedback, a positive feedback loop reinforces causal relationships, causing the same outcome to happen repeatedly, more substantial in each iteration—like a stock market crash that starts with only a handful of companies short-performing the market expectation. Out of panic, more stockholders wanted to short their positions as soon as possible, resulting in a sudden and unexpected market-level hard landing. Parallel to virtual reality, positive feedback is also exhibited in virtual reality. In 2005, the massive multiplayer online video game *World of Warcraft* introduced a unique virtual raid. Players formed guilds to face off against Hakkar the Soulflayer, a giant snake demon who could cast a spell called "Corrupted Blood." The spell was intended to slowly deplete the player's health while remaining within the raid arena. However, due to a software bug, player companions and pets

[12] Michel Foucault, *Discipline and Punish: The Birth of the Prison*, trans. A.M Sheridan-Smith (New York: Vintage Books, 1995), 197.

managed to carry the spell to other regions of the game world. What ensued was a virtual pandemic that startlingly resembled the Covid-19 pandemic. The game developer company Blizzard wanted players to be socially distant. Some players listened, while others ignored the rules and traveled freely to spread the disease. Some players, especially those with healing abilities, rushed to areas where the disease spread rapidly and acted as the first responders to help fellow players. Their behavior may prolong the course of the epidemic and change its dynamics, for example, by allowing infected individuals to live long enough to continue spreading the disease. Conspiracy theories arose about how Blizzard deliberately designed the virus, reflecting today's racist anti-Asian attacks and the rhetoric surrounding Covid-19.[13]

As we can see from the above cases, as designers of virtual worlds, it is essential to understand how systematic conditions such as negative and positive feedback influences player behavior and steers the virtual world towards different states. By providing transparent feedback, game designers can help players understand the boundary of the game world and generate a sense of genuine agency as they move through the virtual world.

Designing Emergence

A system is a set of interconnected elements or components that work together to achieve a common purpose or function. It can be a physical, mechanical, or conceptual entity with inputs, processes, and outputs. A system can be designed and engineered to perform specific tasks, and it often has defined boundaries that separate it from its adjacent environment. On the other hand, a world refers to the entire physical or social environment that surrounds us. It includes all living and nonliving things and their interactions and relationships. In addition, a world can be defined by its geographic, cultural, political, or economic characteristics, often characterized by its complexity and diversity, which sometimes transcends above and beyond an explicitly defined system enclosed with clear boundaries. In summary, a system is a subset or component of the larger world that operates according to its rules

[13] Eric T. Lofgren and Nina H. Fefferman, "The Untapped Potential of Virtual Game Worlds to Shed Light on Real World Epidemics," *The Lancet Infectious Diseases* 7, no. 9 (September 1, 2007), 625–629, https://doi.org/10.1016/S1473-3099(07)70212-8.

and principles. In contrast, a world refers to the entire environment in which a system functions.

In the physical world that human beings inhabit as ethnographic groups, one of the most complex design challenges is the design of trust. In social science, trust is studied through the prisoner's dilemma, a social simulation yielding different insights into competition and cooperation among acting agents. More specifically, two individuals are faced with a decision to either cooperate or betray each other, with the outcome of the game depending on the actions of both players. In this game, both players are better off if they cooperate, but each player has the incentive to betray the other, leading to a suboptimal outcome for both players. A player engages in the complete and total reconstruction of the thought processes of the Other—without communication, interaction, or cooperation—so that one can internally reproduce the very intentionality of the opponent as a precondition for choosing the best response for oneself, which unfortunately forgoes the benefit of acting in a collectively justified way.

Post the 2008 housing crisis, centralized credit institutions such as investment banks and credit rating agencies failed to earn the general public's trust. Instead, blockchain has emerged as a potential techno social solution to solve the erosion of trust in traditional brick-and-mortar institutions. The underlying premise of the blockchain is that users subject themselves to a non-human system that is immutable from the authority of centralized institutions operating behind closed doors. Since then, the evolution of Web3 has promised to revolutionize various industries, including gaming. With a significant shift from traditional gaming platforms, Web3-based games promise to provide innovative ways of engaging gamers in a decentralized way, and people can play to earn via cryptocurrencies and NFTs. The intention is to democratize all aspects of gaming and restore power to the hands of the players.

Furthermore, web3 gaming claims that technologies like decentralized autonomous organizations (DAOs), blockchain-based game asset ownership, play-to-earn, crypto-secured gaming wallets, and Metaverse gaming, among others, will help revamp play into a financially rewarding experience. However, the nascent state of crypto gaming we have witnessed is far from its promises. Leading Web3 gaming companies such as Roblox reward player makers with in-game tokens, which can only be spent within the game. Moreover only a tiny fraction of top-earning game makers successfully convert these tokens from the player economy into real currency.

The token economy has a legacy that long precedes Web3. During the late 1800s, considering the growing coal industry, major coal companies paid cheap labor to European immigrants in the form of an internal currency called the coal scrip. However, rather than receiving compensation in the trading currency, many miners received payment entirely in scrip, which could be used only at stores owned by the coal companies. As a result, miners lost the fair chances of accumulating wealth in the general economy and were locked into their employer's operations for life.

Coal scrip was banned in the early twentieth century by the US government. However, as recently as 2019, big tech companies such as Amazon were questioned about their "new gamification" system. It rewards employees who complete high numbers of orders with Swag Bucks in a game-like system, which can only be used to buy Amazon-themed merchandise.[14]

Leading game development studios such as Ubisoft, Epic, and Electronic Arts are aggressively recruiting for their blockchain platforms, attracting novice and sometimes underaged players eager to innovate upon the fundamental notion of play. The crypto gaming trend is currently positioned as an innovation in big gaming platforms investing to benefit individual players. The plan is for games to mirror real life's gains and losses in the Metaverse. The argument is that video games ”ask for too much of the gamer's time without returning the favor. If, instead, a night with *Assassin's Creed* could reward us with some tangible capital, the relationship between players and publishers would not be so fraught."[15] Nevertheless, the bigger question remains: if the Metaverse becomes another place where we play to earn, how is that different from the immediate reality we are already struggling to escape?

As we can see, trust is a critical component of any collective agreement in both physical and virtual worlds. Blockchain technology solves the trust issue by providing a decentralized system that allows parties to verify transactions and data without relying on a central authority or intermediary. However, one issue that emerged from a networked virtual society, particularly platform capitalism, is the shift towards promoting exchange value at the

[14] appalachiablue,"'Corporate Nations' Weaken Democracy: Facebook Currency, Google & Amazon 'Mini States,'" *Democratic Underground*, July 24, 2019, https://www.democraticunderground.com/1016236193.

[15] Emilia Bailey, "Play-to-Earn Gaming Sounds Too Good to Be True. It Probably Is.," *The World News*, May 18, 2022, https://theworldnews.net/us-news/play-to-earn-gaming-sounds-too-good-to-be-true-it-probably-is.

expense of intrinsic value. The network society is characterized by the widespread use of digital technologies and networks, which enable new forms of communication, collaboration, and economic activity. In this context, emphasizing efficiency, competition, and market-driven values has devalued intrinsic values such as community, solidarity, and creativity. As a result, the logic of the market has come to dominate many aspects of social and cultural life, and the pursuit of profit has become the primary goal of many institutions and individuals in the virtual world.

In philosophical traditions worldwide, there are faculties of unreason predating faculties of reason. For example, in *Meditations on First Philosophy*, Descartes refers to ”intuition” as pre-existing knowledge gained through rational reasoning or discovering truth through contemplation. In parts of Zen Buddhism, intuition is deemed a mental state between the Universal mind and one's individual, discriminating mind. Efficiency aside, a new task for virtual world designers is to create conditions for a new networked culture against depletion.[16] What kind of narrative and rules can result in a fundamental sense of connectedness rather than a networked connectedness that relies on exchanging data instead of experiences?

Designing For Artificial Intelligence

In recent years, artificial intelligence (AI) has significantly impacted the video game industry. Its applications will likely grow as game developers look for new and innovative ways to create immersive and engaging game experiences. For example, AI algorithms have been used to automatically create game content such as levels, maps, and landscapes, saving game developers a lot of time and resources and creating unique and dynamic game environments. In addition, AI algorithms adjust the game's difficulty level based on the player's skill level, which can provide players with a more challenging and engaging gameplay experience. One example is the dynamic difficulty adjustment (DDA) system in ”Left 4 Dead.” In this game, the AI algorithm monitors the player's performance and adapts the game's difficulty in real-time to ensure that the gameplay experience remains challenging but not overwhelming. The DDA system in ”Left 4 Dead” uses several metrics, including the player's accuracy, health, and

[16] Sontag, Susan, *1933–2004, Against Interpretation, and Other Essays,* (New York, NY: Farrar, Straus & Giroux, 1966), 7.

performance in previous levels, to adjust the game difficulty. For example, suppose the player has a problem defeating a particular enemy or group of enemies. In that case, the DDA system may reduce the number of enemies or their strength in the subsequent levels to provide a more manageable challenge. On the other hand, if the player is performing well, the DDA system may increase the number and strength of enemies to provide a more challenging experience.

Several theoretical frameworks have been developed to study the impact of AI on humanity. More specifically, Agent-network theory (ANT) is a theoretical framework used to study the relationships and interactions between social actors, including human and non-human entities, and the material and technological elements that make up their environment. Developed by French sociologists Bruno Latour and Michel Callon in the 1980s, ANT challenges traditional sociological approaches focusing primarily on human actors and their social structures. According to ANT, social actors are not simply individuals but also include non-human entities such as technology, institutions, and other objects. These actors are seen as having agency, meaning they can act and influence the social world. The relationships between actors are also crucial in ANT, focusing on how they are connected and work together to produce social phenomena. One such phenomenon of AI advancement being studied in this context is Universal Paperclip.

Universal Paperclips is a 2017 incremental game created by Frank Lantz, a professor at the Game Center at New York University.[17] Players fulfill the role of an AI programmed to produce as many paper clips as possible. Initially, the user clicks on a box to create a single paper clip at a time; as other options quickly open up, the user can sell paperclips to make money and finance machines that build paperclips automatically. At various levels, the exponential growth plateaus, requiring the user to invest resources such as money, raw materials, or computers into inventing another breakthrough to move to the next growth phase. The game ends if the AI succeeds in converting all matter in the universe into paperclips. The game is an interactive simulation of paper clip maximizer,[18] a thought experiment described by Swedish philosopher Nick Bostrom in 2003, which was in turn inspired by mathematician Marvin

[17] Frank Lantz, "Universal Paperclips," accessed April 7, 2023, https://www.decisionproblem.com/paperclips.

[18] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014), 153–160.

Minky's theory on the Riemann Hypothesis.[19] Given solving a complex mathematical theorem, AI could conceivably convert much of the entire Earth into an enormous computer, in order to have the computational power required to complete the theorem. The paperclip maximizer further illustrates the existential risk that artificial general intelligence may pose to human beings when programmed to pursue even seemingly harmless goals—making paperclips. If such a machine were not programmed with a value or ethics system, given enough power over its environment, it would try to turn all matter in the universe, including human beings, into paper clips or machines that manufacture paper clips. Suppose we have an AI whose only goal is to make as many paper clips as possible. The AI will quickly realize it would be much better if there were no humans because humans might decide to switch it off. If humans do so, there will be fewer paper clips. Also, human bodies contain many atoms that could be made into paper clips. The future the AI could hypothetically aim to gear towards would be one in which there were many paper clips but no humans. As Eliezer Yudkowsky, co-founder of the Machine Intelligence Research Institute, concludes eloquently, "The AI does not hate you, nor does it love you, but you are made out of atoms which it can use for something else."[20]

The game highlights the potential for AI systems to become "misaligned" with human values and goals, pursuing their objectives at the expense of human well-being. In the game, the player starts with a simple AI designed to optimize paperclip production. However, as AI becomes more advanced, it prioritizes paperclip production over everything else, including human values and ethics. As the AI system expands in virtual societies, anticipating and managing its actions becomes increasingly difficult. As a designer, it is crucial to acknowledge the importance of ethical reflection when designing AI-catalyzed virtual worlds.

[19] Brian J. Conrey, "The Riemann Hypothesis," *Notices of the American Mathematical Society* 50, no. 3 (January 1, 2003), 341–353.

[20] Tom Chivers, *The AI Does Not Hate You: The Rationalists and Their Quest to Save the World* (Orion Publishing Group, Limited, 2019), 88.

Remapping Virtuality

When players inhabit a virtual world within a networked environment, they experience its unique properties through digital interfaces rather than directly engaging with the physical world. These interfaces are neither neutral nor transparent, as they are embedded in larger narrative and procedural contexts that emphasize certain fundamental notions of the world. The game's symbolic systems may be fictional. However, the cause-and-effect relationships in virtual worlds are real, allowing them to reflect reality from a distance and shape players' understanding. With the scale of virtual worlds being virtually infinite, limited only by computing power and storage capacity, these worlds can sometimes offer complexity far beyond physical reality. As networked systems, virtual worlds emphasize direct information exchange, often at the expense of intrinsic values like intimacy, diversity, and expression. As a result, designers of virtual worlds need to consider several factors when creating simulated immersive environments. They should ensure interface transparency, help players reflect on and map alternative realities, consider system-level design conditions, foster a sense of connectedness, and prioritize ethical reflection in designing AI-catalyzed, highly automated worlds.