

Alice Leteissier

Affective Scripts: Gesture, Atmosphere, and Narrative Collapse in User-Generated Short-form Videos

This written component is part of an MFA thesis project at the Academy of Fine Arts, University of the Arts, Helsinki.

Date of submission of the thesis project : 2nd of April 2026



Abstract

This thesis consists of a written component and an artistic project developed in dialogue. The artistic component, *Signaling* is a one-minute looping video installation projected in a black box. Drawing on post-apocalyptic cinema and mainstream trailer aesthetics, it constructs an emotionally saturated atmosphere that resembles narrative without offering resolution. Through repetition, interruption, and the use of archetypal figures, the work simulates legibility while destabilizing affective investment. The written component takes *Signaling* as its point of departure and conceptual engine. In dialogue with affect theory, atmosphere theory, and media studies, it examines how user-generated short-form video reorganizes attention through gesture, repetition, and mood. Emphasis is placed on the reciprocal relationship between practice and theory, with the artwork functioning as both method and object of inquiry.

Artistic component : ***Signaling*** (2025), video installation, 4K, 16:9, 1min.
(presented at Kuvan Kevät 2025 at the Academy of Fine Arts Helsinki)

Supervisor : Salla Tykkä

Examiners : Itha O'Neill, Juha-Heikki Tihinen

Table of Contents

Introduction

1. Signaling as Method: Installation, Parafiction, and Affective Dispositif.....5
2. Affective Infrastructures: User-Generated Video and Post-Narrative Media.....6

1. Embodying Narratives: Main Character Energy 8

- 1.1 Narrative Compression and the Collapse of Plot 8
- 1.2 What Does the Main Character Do? 11
- 1.3 User-Generated Short-Form Video as Affective Script 14

2. The Afterlife of Gesture: The NPC and the Vibe 19

- 2.1 The NPC as Cultural Figure 19
- 2.2 Gesture Without Intention 21
- 2.3 Loops, Repetition, and Cybernetic Shame 23
- 2.4 From Repeated Gesture to Atmosphere 25

Synthesis 28

Bibliography 29

A generative AI tool (ChatGPT) was used in a limited capacity to support language refinement. All substantive arguments, analysis, and conclusions were developed by the author.

Introduction

1. Signaling as Method: Installation, Parafiction, and Affective Dispositif

This thesis emerges from an artistic practice concerned with the affective conditions of contemporary image culture. At its center is *Signaling*, my MFA exhibition work and the conceptual engine of this inquiry. Presented as a one-minute looping video installation in a black box, the work adopts the grammar of post-apocalyptic cinema and mainstream trailer aesthetics: dramatic music, cinematic color grading, archetypal figures, and smooth camera movements. Yet it offers no resolution. The video runs on a closed loop, punctuated by thirty seconds of black screen and ambient sound. Viewers enter an emotionally saturated atmosphere that resembles narrative but refuses to unfold.

The work compresses recognizable tropes, simulates legibility, and interrupts affective investment before it can stabilize. It asks what happens when image and atmosphere override plot, speech, and sequence. In this sense, *Signaling* functions as a mimetic-critical experiment: it reproduces the affective logic of contemporary media in order to expose its operations.

This thesis develops out of questions that arose during the making of *Signaling*. Rather than resolving these questions, it treats writing as a parallel site of inquiry—one that allows them to be unfolded, tested, and reframed. It asks how user-generated short-form video, through repetition, compression, and sensory saturation, cultivates a post-narrative sensibility—one that privileges archetypes over characters, gestures over speech, and atmospheres over plots.

The research developed in this thesis directly informed my artistic practice by providing a framework through which to understand and manipulate the affective logics of contemporary image circulation. Rather than opposing these logics, my work adopts and intensifies them—compressing narrative, foregrounding gesture, and structuring viewing through repetition—in order to expose their underlying mechanisms.

Indeed, by staging what appears to be a plausible trailer for a non-existent film within a gallery installation, the work operates in proximity to what Carrie Lambert-Beatty defines as parafiction¹: a practice that mobilizes the codes of reality to produce something that hovers between the credible and the constructed. While *Signaling* does not construct a fully coherent fictional world, it borrows the aesthetic authority and affective immediacy of cinematic language to simulate narrative legibility.

Thereby, rather than simply challenging the motif of the long-duration museum video piece, the decision to present a one-minute trailer within the black box of a gallery functions as both a performative gesture and an experimental dispositif. It produces a subtle disorientation by confronting the viewer with a familiar audiovisual form displaced into a

¹ Carrie Lambert-Beatty, "Make-Believe: Parafiction and Plausibility," *October* 129 (2009)

context where a different temporal and formal experience is typically expected. This installation format therefore positions the viewer within the same affective conditions the work seeks to analyze.

This approach is informed by contemporary artists working at the intersection of video, performance, and technological systems, whose practices explore the instability of bodily expression and the mediation of affect. Ryan Trecartin stages hyper-mediated subjectivities through fragmented, accelerated narratives that blend performance, digital simulation, and social media aesthetics. Cécile B. Evans produces affect-laden narratives by combining found footage, CGI, 3D animation, and anime, examining the emotional and corporeal experience of digital environments. Ed Atkins highlights the disjunctions between voice, image, and digital bodies, while Ana Estériola's uncanny installations, such as *Reproduced*², experiment with fragmented speech and absurd improvised movie dubbing. Across these practices, shared concerns emerge: the gap between verbal expression and physical embodiment, and the ways affect and subjectivity are constructed, mediated, and destabilized within technological systems. *Signaling* contributes to this field by translating these logics into short-form, spatial, and cinematic formats, using compression, repetition, and immersive audiovisual strategies to probe how certain codifications of affect detach from their narrative origins and begin to circulate as autonomous formulas.

2. Affective Infrastructures: User-Generated Video and Post-Narrative Media

User-generated short-form video has become one of the defining media formats of contemporary digital culture. Circulating endlessly across algorithmic platforms, these videos rarely rely on narrative development, instead engaging viewers through compressed emotional cues: body language, gestures, tropes, atmospheres, and nonverbal affect. This thesis examines how User-generated short-form video—produced and circulated within ecosystems such as TikTok, Instagram Reels, and YouTube Shorts—reorganizes perceptual and affective orientation by privileging sensation over sequence, recognition over resolution, and mood over plot.

Grounded in artistic practice and image-based thinking, and in dialogue with media and affect theory, I argue that this content operates less as a narrative medium than as an affective infrastructure. Rather than organizing experience primarily through narrative progression, it produces conditions of feeling. These conditions—often colloquially described as *vibe*—function as immersive environments that modulate attention, identification, and desire without requiring narrative comprehension.

Following thinkers of affect theory and atmosphere theory such as Gernot Böhme³, and Eve Kosofsky Sedgwick⁴, this thesis approaches contemporary media not as a sequence of discrete messages, but as a system of ambient affective environments. Images

² Ana Estériola, *Reproduced*, video installation, 2017.

³ Gernot Böhme, *The Aesthetics of Atmospheres*, ed. Jean-Paul Thibaud (New York: Routledge, 2016).

⁴ Eve Kosofsky Sedgwick, *Touching Feeling: Affect, Pedagogy, Performativity* (Durham: Duke University Press, 2003), 8–10.

are no longer encountered one by one; they form immersive fields. Content becomes mood. The vibe becomes the message.

Signaling literalizes this condition. Its trailer-like syntax draws viewers in through familiarity, only to strand them in repetition. The recurring black screen operates as an affective rupture—a break in continuity that recalls Eve Kosofsky Sedgwick and Adam Frank understanding of shame as interruption⁵. Viewers report confusion, anxiety, or a compulsive desire for resolution. Some disengage; others remain, scanning for a resolution that never arrives. This affective friction is not incidental. It mirrors the way short-form video produces intensity without consequence, emotion without narrative transformation.

Across two chapters, I trace how narrative does not disappear but mutates, persisting as *affective trace*: first through the figure of the Main character, and then through its negation in the NPC (Non-player character) and the emergence of vibe as an environmental condition.

Together, these chapters investigate how user-generated short-form video may be understood as cultivating a post-narrative sensibility—one in which affect is carried predominantly by gesture, repetition, and atmosphere.

⁵ Eve Kosofsky Sedgwick and Adam Frank, “Shame in the Cybernetic Fold,” in *Touching Feeling: Affect, Pedagogy, Performativity* (Durham, NC: Duke University Press, 2003)

1 — Embodying Narratives : *Main Character Energy*

1.1 Narrative Compression and the Collapse of Plot

Contemporary criticism often approaches new media by declaring the exhaustion or irrelevance of narrative. Linear storytelling is dismissed as too Western, too teleological, too invested in coherence and closure⁶. Across game studies and digital media theory, narrative is frequently replaced by concepts such as interaction, simulation, or play⁷. As McKenzie Wark and others have argued, digital culture is better understood through the logic of games than through that of stories—an idea echoed in McKenzie Wark’s *Gamer theory*’s claim that there is nothing outside the game space : “Play no longer functions as a foil for a critical theory. The utopian dream of liberating play from the game, of pure play beyond the game, merely opened the way for the extension of gamespace into every aspect of everyday life.”⁸

This critical shift is not without reason. Canonical theories of narrative, such as Frank Kermode’s account of emplotment and endings⁹, have been rigorously challenged by postmodern, queer, and postcolonial thinkers. Scholars such as Elizabeth Freeman have emphasized temporalities that resist linear progression—queer time, suspended time, cyclical time—undermining the universality of narrative closure. Narrative, in this view, becomes suspect not because it organizes experience, but because it enforces normative structures of meaning.

Yet to reject narrative altogether is to overlook its persistence. Narrative desire does not disappear; it mutates. Even when plot collapses, the expectation of narrative remains. What changes is its scale and mode of operation.

In short-form video, this mutation takes the form of compression. Narrative does not unfold through extended arcs of causality and consequence, but is condensed into micro-gestures, atmospheres, and affective cues. The question, then, is not whether narrative still exists, but how it survives under conditions of speed, repetition, and algorithmic circulation.

The concept of *Main character energy* offers a revealing entry point into this transformation. Emerging from meme culture and gaining widespread popularity on platforms such as TikTok and Twitter between 2019 and 2021, main character energy names a way of inhabiting the world as if one were the center of a story. To “have” main character energy is to appear confident, intentional, and narratively privileged—as though events unfold in relation to oneself. Significantly, this energy is described not as an action or achievement, but as an aura: something perceived, projected, and recognized by others.

What, then, is being performed when someone enacts main character energy? If narrative is no longer sustained through extended plots, what takes its place? The answer is not interaction in the literal sense, nor the logic of games alone. Rather, user-generated

⁶ Elizabeth Freeman, *Time Binds: Queer Temporalities, Queer Histories* (Durham: Duke University Press, 2010),

⁷ McKenzie Wark, *Gamer Theory* (Cambridge, MA: Harvard University Press, 2007),

⁸ *Ibid.*, 19.

⁹ Frank Kermode, *The Sense of an Ending: Studies in the Theory of Fiction* (Oxford: Oxford University Press, 1967).

short-form video stages a form of narrative role-play that remains suspended rather than developed. Users inhabit recognizable narrative positions—protagonist, observer, disruptor—through gesture, framing, and affect, allowing these roles to be legible without requiring progression or closure.

In this sense, user-generated short-form video resembles a form of everyday LARPing (live-action role-playing), not of fictional characters, but of narrative roles themselves. There is no game without rules, and here the rules are inherited from narrative convention: centrality, legibility, and affective resonance. What disappears is not story, but plot—the structured sequence of actions that produces irreversible change.



Fig. 1. *Signaling* (2025), video installation, HD video still showing a character with white eyes, staring at its own reflection in a mirror, 16:9, 1 min loop. Image courtesy of the artist.



Fig. 2. *Signaling* (2025), video installation, HD video still showing a character with a tattoo reading “Bad Boy” on the skull, standing in a white void, 16:9, 1 min loop. Image courtesy of the artist.



Fig. 3. *Signaling* (2025), video installation, HD video still showing characters crouching in a dimly lit warehouse, 16:9, 1 min loop. Image courtesy of the artist.



Fig. 4. *Signaling* (2025), video installation, HD video still showing two characters kneeling side by side. Image courtesy of the artist.

Instant character legibility is precisely what *Signaling* operates on and intensifies. The installation mobilizes a compressed narrative grammar drawn from cinematic and gaming archetypes, relying on immediate recognition rather than development. A Trinity-like figure moves slowly through a dimly lit space, theatrically licking a knife. A naked cyborg stands isolated in a sterile white void, the words *Bad Boy* tattooed across his skull. Motionless bodies lie scattered across public spaces, asleep or unconscious. Within seconds, the viewer knows how to read these figures: hero, threat, victim, ally. Narrative positions are assigned instantly, without explanation.

Crucially, none of the characters speak. There is no dialogue, no exposition, no causal chain linking one image to the next. Instead, bodies crouch, glance around suspiciously, chase one another, or turn abruptly toward the camera. Gesture, posture, costume, lighting, and framing do the narrative work. Meaning is not constructed through sequence, but through archetypal aura. The figures appear saturated with narrative potential, yet that potential is never actualized.

Signaling thus stages narrative compression as a formal condition : It offers the promise of story without its fulfillment, inviting recognition while withholding progression. Like short-form video, it relies on familiar narrative codes only to suspend them in a loop. What the viewer encounters is not a story unfolding in time, but a sequence of affective cues that remain perpetually on the verge of significance.

This chapter goes on to examine *Main character energy* as the site where narrative compression becomes visible. By analyzing how classical storytelling structures are reduced to gestures, poses, and moods, it argues that short-form video replaces plot with micro-affect. Narrative survives not as sequence, but as sensation; not as causality, but as legibility. The main character no longer drives the story forward—time simply passes through them.

1.2. What Does the Main Character Do?

User-generated short-form videos are saturated with main characters, yet remarkably poor in events. People wake up, sit, stare, stretch, apply skincare, or do nothing at all. And still, these videos function. They attract attention, identification, and projection. This suggests that the logic of the Main character in short-form video is not based on action or plot, but on the legibility of presence.

Body language seems to be one of the most efficient vehicles for conveying affect within the compressed temporal economy of short-form video. Human attention is disproportionately drawn to other human bodies—especially faces— through deeply ingrained perceptual and social attunement. Research across multimedia learning, social psychology, and human–computer interaction suggests that information presented through a visible human presence tends to elicit greater engagement, trust, and retention than text alone.¹⁰ Unsurprisingly, user-generated video is dominated by scenes of creators talking, dancing, cooking, unboxing products, performing pranks, or exercising. Yet the specific activity matters remarkably little. What sustains attention is not action in the narrative sense, but the legibility of embodied presence itself. In this context, the main character does not need to *do* anything; appearing to be absorbed, intentional, or affectively available is sufficient.

¹⁰ Richard E. Mayer, *Multimedia Learning*, 2nd ed. (Cambridge: Cambridge University Press, 2009); Roxana Moreno and Richard E. Mayer, “Engaging Students in Active Learning: The Case for Personalized Multimedia Messages,” *Journal of Educational Psychology* 92, no. 4 (2000); Charlotte N. Gunawardena and Frank J. Zittle, “Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment,” *American Journal of Distance Education* 11, (1997)

In classical storytelling the Main character is distinct from the protagonist, whose actions and desire drive the plot and the action. Even if in some cases they can be one and the same, the Main character is first and foremost the one through which we experience the story. The plot unfolds in relation to them, and an exterior event comes to matter as long as it affects them.

So what are the signs that you are watching a Main character? A main character establishes a position around which attention can organize itself. Whether they live alone in a cabin in the woods, work as a scuba diving instructor, or perform the figure of the independent businesswoman, the lifestyle on display functions less as an object of desire than as a stable point of orientation. One does not need to want this life, or even comprehend it fully. The performance of the main character within their environment is sufficient to signal that things revolve around them—that their presence anchors the scene. This anchoring invites projection without requiring aspiration.



Fig. 5. Still from a TikTok video by @miayilin (2025). Creator seated at a desk in a large library, reading and taking notes on a tablet.

Fig. 6. Still from a TikTok video by @emmahoulmiere3 (2025). Creator seated at the back of a motorboat, hair blowing in the wind, looking toward the sunset.



Fig. 7. Still from a TikTok video by @ingridohara (2025). Creator standing barefoot in a towel on a street, with bystanders visible in the background.

To understand the Main Character’s role, we can examine specific performances. What do their actions signal, and how do they structure the viewer’s attention? The examples below reveal how MC performances convey presence, focus, and affect in short-form video:

(Fig 5) The Main character can be seen sitting at a desk in a bright spacious luxurious library. Their expression is that of deep concentration, reading a book and taking notes on her tablet. Never acknowledging the Camera. On top of the video is a face caption that reads “study hard. workout. practice self-care. eat healthy. chase your dreams.”

(Fig 6) In this other video, MC is sitting at the back of a motor boat driving at full speed. MC’s long blond hair is flowing in the wind as she looks at the sunset on the horizon. Caption “It will take me maybe 2 years, maybe 5, but I will do whatever it takes to accomplish what I promised myself”.

In these last two examples, the Main character is a focused surface, that appears absorbed, intentional, or self-contained but does not reveal inner thought, conflict, or desire functions primarily as a legible affective signal. In classical narrative focus usually implies

goal-directed action : desire, conflict, future consequence whereas in short-form video, focus becomes pure appearance, a visual pose of intentionality.

The MC also acts as a temporal anchor: time passes through them, not because of them. Despite all the activities depicted, the sense of productivity is less about accomplishment than about maintaining an aura of control. This explains the emphasis on repetition disguised as “routines” and the importance of appearance. In another video by @miayilin (**Fig. 5**), the MC performs several consecutive actions: brushing teeth and applying face cream, typing on a computer in a state of deep focus, running on a treadmill, eating low-fat yogurt and fruit for breakfast, applying more skincare and make-up, and eating a salad while reading.

Finally, the Main character performs disruption without consequence. In classical narrative logic, disruption is known to create conflict, for instance by breaking a social norm. Most importantly it produces consequences and forces transformation (for the protagonist or the world). But in short-form video, disruption is permitted only insofar as it remains legible, entertaining, and consequence-free, as illustrated in the following example : a scandalous *Get ready with me* video :

(Fig 7) People wait in line for a seemingly classy event. Suddenly, MC appears—barefoot, wrapped in a towel, dripping wet, her face covered by a white mask, carrying a cardboard box. She briefly hides behind a car, applies facial products, and receives help from a bodyguard to rinse her hair. As the crowd watches and films, she dresses in a long, shiny pink gown, and checks her reflection in a car mirror. After the final touch—pink glossy lips—she walks past the line and through the event doors, leaving the crowd standing in awe of her sudden, striking beauty.

The disruptive act stays in the realm of pure visibility. Is framed aesthetically, contained within the clip and produces no aftermath. The crowd observes but does not intervene. The bodyguard assists. The social order remains intact. What looks like rebellion is actually licensed spectacle.

These gestures—studying, standing, resting, preparing—are not actions in a narrative sense. They are affective markers: traces of intention. They allow the viewer to recognize a state, project a future, and imagine continuity without requiring resolution. In the next section, I argue that short-form video is not merely compatible with this kind of main character energy, but structurally optimized for it.

1.3. User-Generated Short-Form Video as Affective Script

In User-generated short-form video the human body becomes a narrative engine. Rather than relying on plot development or dialogue, these videos operate through micro-gestures: fleeting glances, subtle facial movements, poses, and camera-facing confessions that instantly signal archetypal roles recognizable within platform culture. What circulates under the label of *Main character energy* is not a stable identity or psychological state, but a gestural code—a bodily rhetoric shaped by algorithmic visibility, aesthetic conventions, and the pressures of an attention economy. Within this economy, the Main

character's pose emerges less as self-expression than as a survival strategy: a way of becoming legible, memorable, and repeatable within an endless feed.

Why, then, does the ten-to-fifteen-second user-generated video appear as the ideal site for the expression of Main character energy? This question cannot be answered by appealing solely to duration. To frame short-form video as “short” is already misleading. Such a definition relies on an attention-span discourse that treats time as the primary variable of meaning. Instead, this chapter approaches short-form video from another angle: not as a temporal format, but as a vehicle for producing what I call an *affective trace*—a residue of feeling, recognition, or identification that persists beyond the moment of viewing.

Within a traditional media theory, User-generated short-form video might be approached as a format distinguished primarily by the conditions of its creation and distribution on algorithmic platforms. Yet the algorithms that govern visibility, recommendation, and circulation remain largely opaque, even to those who work inside these companies. As accounts from TikTok and Douyin employees suggest¹¹, engineers themselves often lack a comprehensive understanding of how these systems operate, as knowledge is fragmented across teams and hierarchies. The algorithm thus functions as a black box, concentrating power while remaining structurally unreadable. In this context, creators cannot meaningfully address the algorithm directly. What remains available for interpretation and manipulation are aesthetic codes: gestures, tropes, formats, and performative conventions. These become the de facto interface through which creators attempt to negotiate visibility.

Art historical approaches might be attempting to situate user-generated content by comparing it to existing formats—film, photography, experimental video, or self-portraiture. Such comparisons frequently return to the issue of duration, framing short-form video as a symptom of diminished attention spans and cultural decline, in contrast to the supposed depth of long-form media. Claire Bishop has identified this tendency as part of a broader moralization of attention¹². Across discourses on cinema, television, and social media, deep attention is repeatedly associated with agency, autonomy, and even humanity itself, while distraction becomes a term of accusation. Distraction, Bishop argues, is not the opposite of attention but a judgment about misplaced attention, often mobilized by dominant social groups to discredit others. It implies a weakness of character rather than acknowledging structural conditions of perception.

In her book *Disordered attention*, Claire Bishop also cites N. Katherine Hayles's concept of *hyperattention*, characterized by rapid task-switching, multiple information streams, and a low tolerance for boredom. While often framed as antithetical to the deep attention valued in the humanities, Hayles argues that hyperattention is not inherently deficient but culturally adaptive. Yet older patterns of thinking persist. For theorists such as Bernard Stiegler, hyperattention signals dehumanization, a loss of consciousness likened to animal perception. These debates reveal that attention is not a neutral cognitive capacity but

¹¹ Chunmei Su, *Douyin, TikTok and China's Online Screen Industry: The Rise of Short Video Platforms* (New York: Routledge, 2022)

¹² Claire Bishop, *Disordered Attention: How We Look at Art and Performance Today* (London: Verso, 2024).

a socially constructed value system. The binary between attention and distraction obscures the ways in which different media demand different modes of engagement.

From this perspective, the distinction between time-based and still formats appears increasingly inadequate. What defines a format today is not the amount of time it occupies, nor the duration of engagement it prescribes, but the kind of attention it solicits. Hito Steyerl's analysis of contemporary video art is instructive here. In *Is the Museum a Factory?*, she observes that much contemporary video is produced with the assumption that viewers will encounter it fragmentarily—watching portions, entering and exiting at will, yet still grasping its meaning¹³. Partial viewing is not a failure of reception but a structural condition of the work itself. User-generated short-form videos operate according to a similar logic: they are designed to withstand interruption, repetition, and incomplete attention without losing legibility.

Paradoxically, many user-generated short videos share more in common with photography than with cinema. A significant proportion consists of a largely static face, minimal movement, and an overlay of text or music. A typical example is the ubiquitous genre of videos showing a close up of a woman facing a camera accompanied by a melancholic pop song and a lengthy caption on top of the video. Why is this a video rather than a photograph? The answer lies in what the video preserves: the blinking of eyelashes, the slight twitch of the mouth, the shifting shadow across the forehead. These micro-movements constitute the affective core of the clip. The viewer's investment is in the attentiveness to minute bodily changes.

To grasp this dynamic, it can be useful to consider a very different medium: early Renaissance painting. *The Tribute Money* by Masaccio serves as a canonical example of continuous narrative within a broader Western tradition that, from antiquity onward, organized multiple temporal moments within a single image. This comparison is not intended to equate these media, but to place them on an equal analytical footing as dominant image regimes of their respective periods, and to clarify how different historical conditions organize the relationship between gesture, time, and narrative. In *The Tribute Money*, Masaccio compresses multiple narrative moments into a single visual field, allowing several temporal instances to coexist within one image. These moments are spatially distributed but remain narratively linked: the viewer reconstructs a sequence by scanning the composition, moving between gestures that belong to different points in time but form a coherent story. For a fifteenth-century viewer, this encounter would typically be sustained and site-specific, inviting prolonged contemplation in which gestures are gradually internalized as part of a unified narrative whole.

By contrast, short-form video isolates gestures rather than embedding them within a continuous sequence. Instead of multiple moments coexisting within a single image, each clip presents a fragment detached from broader narrative context. These fragments are consumed discontinuously—glimpsed, interrupted, and revisited across the day—within an endless, algorithmically sequenced feed. Where Renaissance painting condenses time into a stable composition that can be repeatedly examined, short-form video disperses time across a series of fleeting encounters. Affective meaning no longer emerges from

¹³ Hito Steyerl, "Is a Museum a Factory?" in *The Wretched of the Screen* (Berlin: Sternberg Press, 2012)

reconstructing a narrative within a single work, but from the accumulation of decontextualized gestures encountered across many clips. The question then becomes not how a story unfolds within an image, but what kinds of movement, sensation, and latent narratives emerge from this continuous flow of isolated, repeatable gestures.



(from left to right)

Fig. 8. Masaccio, *The Tribute Money* (c. 1425), detail: Saint Peter extracting the coin from the fish.

Fig. 9. Masaccio, *The Tribute Money* (c. 1425), detail: Saint Peter confronting Christ and the tax collector.

Fig. 10. Masaccio, *The Tribute Money* (c. 1425), detail: Saint Peter paying the tax collector.

From the perspective of production, the most significant difference is that in user-generated content, the main character is also the director. The subject frames their own gestures, edits their performance, and determines the narrative cues it will communicate. This collapse of roles aligns closely with Erving Goffman's¹⁴ analysis of the presentation of self as a series of performances calibrated for an audience. However, in short-form video, the audience is abstracted and mediated by algorithmic distribution. This dynamic can be further understood through Silvan Tomkins' script theory, that defines scripts as patterned sequences through which affects are organized and repeated over time¹⁵: short-form video offers ready-made affective scripts—sadness, confidence, transformation, vulnerability—that can be rapidly activated, repeated, and modified. The logic of rupture and repetition

¹⁴ Erving Goffman, *The Presentation of Self in Everyday Life* (London: Penguin Books, 1990.).

¹⁵ Silvan S. Tomkins, *Affect Imagery Consciousness*, vol. 3: *The Negative Affects: Anger and Fear* (New York: Springer, 1991)

embedded in the format allows subjects to rehearse versions of themselves continuously, without direct consequence.

This dynamic is particularly visible in videos that stage moments of apparent transgression, such as towel-clad street makeovers or public displays of impropriety. These acts traditionally signal narrative rupture: a break from social norms, an invitation to conflict or transformation. Yet within the short-form video frame, the disruption is immediately neutralized. The surrounding world absorbs it without resistance; the crowd waits, the bodyguard assists, the event proceeds. What appears as rebellion functions instead as spectacle. Main character energy here depends on transgression that is instantly resolved through visibility rather than confrontation.

Ultimately, what viewers seek in Main character performances remains consistent with inherited narrative desires: rupture, transformation, and drama. What has changed is the scale and consequence of these experiences. Short-form video offers them in miniaturized, loopable doses that never alter anything. The appearance of becoming the protagonist substitutes for narrative stakes, and affective traces replace lasting transformation. In this sense, short-form video emerges as the optimal vehicle for *Main character energy* precisely because it allows exceptionalism to be performed endlessly, without accountability.

2 — The Afterlife of Gesture: the NPC and the vibe

2.1 The NPC as Cultural Figure

The term *non-player character* (NPC) originates in tabletop role-playing games, where it designates characters controlled by the game master rather than by players. In digital games, NPCs are computer-controlled entities governed by predetermined scripts and behavioral rules. Unlike player characters, NPCs do not possess agency in the narrative sense; they exist to populate the world, trigger interactions, or facilitate progression. Their actions are automatic, reactive, and constrained, unfolding in response to external inputs rather than internal desire. While NPCs may appear lifelike or expressive, their behavior is fundamentally procedural. They do not initiate stories; they sustain the environment in which stories occur.

From this technical definition emerged a cultural metaphor. Beginning around 2018, *NPC* became an online insult, used to describe individuals perceived as lacking independent thought or agency. This usage often circulated through memes such as the grey-faced, expressionless Wojak, depicting the NPC as emotionally flat, cognitively empty, and mechanically responsive. The insult relies on a moral hierarchy of subjectivity: to be human is to be intentional, reflective, and self-directed, while to be an NPC is to merely react, repeat, and comply. In this sense, the NPC functions as the negative mirror of the main character. Where the main character embodies narrative centrality, uniqueness, and exceptionality, the NPC represents interchangeability, automation, and narrative redundancy.—a figure designed not to be remembered, but to be encountered repeatedly without consequence.

This metaphor has recently been literalized through the rise of NPC livestreaming on platforms such as TikTok. In this format, creators perform as NPCs in real time, responding to viewer inputs—often in the form of paid digital “gifts”—with scripted gestures, phrases, or facial expressions. Viewers effectively assume the role of puppeteers, triggering the performer’s actions through micro-transactions. Popularized in mid-2023, NPC streaming has proven highly lucrative, with creators such as Pinkydoll gaining hundreds of thousands of followers and earning substantial daily income. Here, the NPC is no longer a passive insult but an active performance strategy. Yet the logic remains the same: the performer relinquishes narrative agency in favor of responsiveness, transforming the self into a programmable interface. In contrast to main character energy, which simulates autonomy and exceptionality, NPC performance stages the pleasure of repetition, legibility, and control—revealing the other pole of subjectivity in short-form video culture.

If the Main character centers gesture as expressive self-staging, the NPC marks the moment when gesture detaches from intention and becomes repeatable, legible, and operational. Gesture no longer appears as the outward manifestation of an inner state, but as a surface action calibrated for recognition and response. In this sense, gesture ceases to belong fully to a subject and instead becomes available to a system.

This figure has also informed my own practice, in which NPC-like behaviors are explored within scripted narrative scenes as a way of testing the limits of this cultural metaphor beyond its gaming origins. Although the NPC is not an established concept in

narrative theory or film studies—despite parallels with background characters, stock figures, or flat characters—I found its analytical potency to emerge most clearly in the context of actor direction.

Classical acting techniques are traditionally grounded in mimesis: the production of a coherent world through the alignment of a character’s perceived interiority and outward bodily expression. Gesture, in this framework, functions as the visible trace of intention. In my own video works, such as *People can’t stop talking about crab*, my experiments disrupt this alignment by asking performers to execute highly coded gestural repertoires—those of a cold blooded killer or a bored child—while withholding or misaligning the corresponding intention. The resulting performances produce a sense of dissonance and awkwardness, oscillating between the disturbing and the comic.

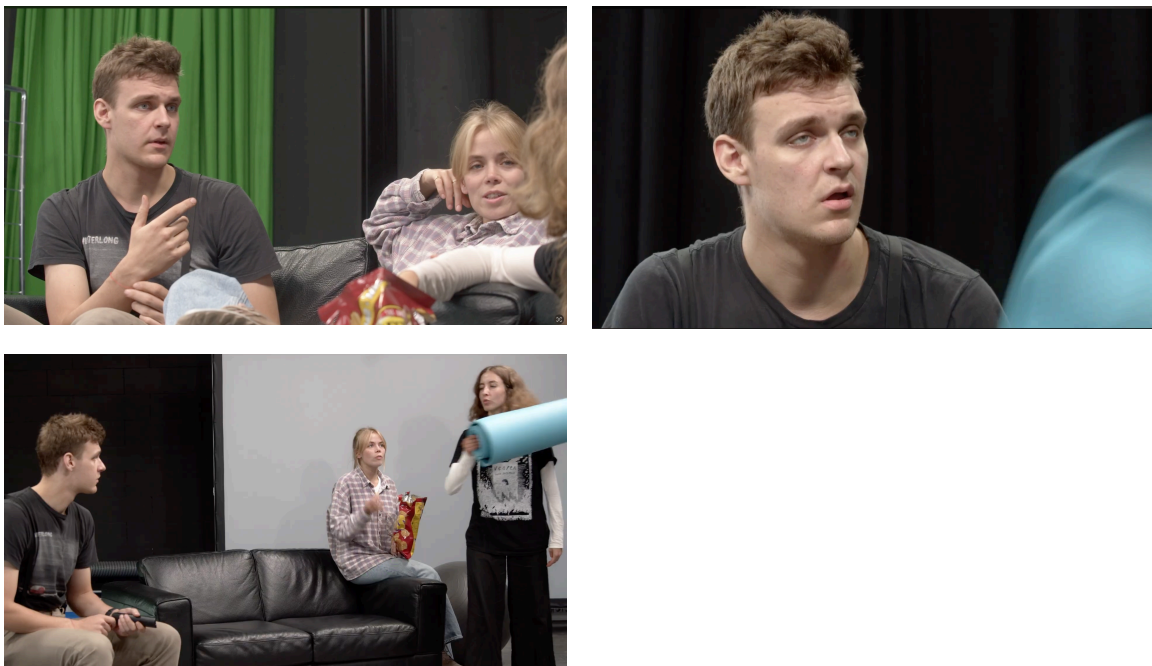


Fig. 11. Fig. 12. Fig. 13. *People can't stop talking about crabs* (2024), video installation, HD video stills demonstrating NPC behaviours in a scene of conflict , 16:9, 1.30 min loop. Image courtesy of the artist.

While inspired by the ecstatic, hyper-expressive characters of Ryan Trecartin’s *I-Be Area*, my previous video works departed from this lineage by foregrounding not excess but restraint. Rather than embodying the frenetic energy of early internet performance, these characters appear resigned, tentative, and uncertain, gesturing in search of connection while no longer trusting the expressive capacity of their own bodies.

In *Signaling*, this process is reversed. Characters are constructed to appear deeply intentional using minimal narrative means, exposing how perceived intentionality itself relies on inherited cultural codes and narrative conventions. A single gesture—such as a fearful gaze directed toward the sky—invites the viewer to project a story. Yet when the story fails to unfold and the gesture returns in a loop, its status begins to shift: from narrative cue to

patterned signal, from expressive act to aesthetic object. What remains is a simulacrum of intention sustained through repetition.

This tension between gesture as expression and gesture as signal forms the basis of the following section.

2.2 Gesture Without Intention

A considerable part of user-generated short-form video falls into the category of life hacks videos : silent videos consisting of gestures showcasing more efficient techniques to perform everyday tasks, such as cutting vegetables, putting on socks, or drinking orange juice. Some of these demonstrate creative and unexpected ways to perform tasks in order to be more efficient, while some introduce a new tool that is supposed to facilitate the completion of a given task. For instance one of these videos showcases a new technique for opening a walnut with some water and a plastic box. But these videos do not merely propose solutions; they standardize gestures, extracting them from context and circulating them as universally legible movements.

As a response to these, a new trend emerged on platforms, consisting in taking these life hacks to extreme extents, by imagining the most absurd solution to the most absurd problems. One of these videos explains how to untangle one's phone charger from a coffee mug and a pair of pliers.

Kaby Lame, the now famous Senegalese-italien Tiktok influencer, became successful by creating content that mirrored these absurd life hacks videos, juxtaposing them with shots of himself solving these same tasks in the most usual and unsurprising manner. In one of his videos, he is reacting to a trend consisting of opening a Cola can with a can opener. We can see him sitting at a table, demonstrating how one can easily open the cola can using the metallic piece at the top of the can, designed for this effect.

All of Kaby's lame videos follow the same format. First we see the original life hack video he is reacting to, these videos are often anonymous, showing only the hands of the person doing the action. This is followed by a still shot of himself facing the camera with a deadpan expression, while performing the same task in the most simple and unbothered manner. Once he is done he takes a small pause, stares incredulously in the camera, and does this gesture of shrugging with his hands to the side. The main point here is that Kaby Lame does not perform as a character, nor does he narrate. His body operates as a corrective mechanism, reasserting obvious gestures without commentary. His shrug is not an expression of disbelief, but a terminal gesture—closing the loop rather than opening interpretation.



(from left to right)

Fig. 14. Still from a TikTok video by @khaby.lame, showing the original life-hack video being re-enacted (TikTok, 2025).

Fig. 15. Still from a TikTok video by @khaby.lame, showing Khaby Lame opening a Coca-Cola can (TikTok, 2025).

Fig. 16. Still from a TikTok video by @khaby.lame, showing Khaby Lame doing his iconic shrugging and hand gesture (TikTok, 2025).

Aby Warburg's concept of Pathosformel provides an interesting tool to approach this new gestural language developing in the ecosystem of short video platforms. Pathosformel designates a "formula of pathos" or emotionally charged gesture that survives across history as a recognizable expressive unit¹⁶. Warburg was concerned with the survival and the afterlife of gestures, which he studied through a lifelong iconographic research that resulted in the *Mnemosyme Atlas*. He tried to understand how similar human gestures reappeared throughout different periods, sometimes expressing opposite affects. His idea being that gesture persists not because it is meaningful to a subject, but because it is remembered by culture. In the ecosystem of short video platforms, we might observe a parallel phenomenon: gestures circulate through algorithmic systems that reproduce, amplify, and recontextualize them across feeds. In this sense, gestures acquire a kind of digital afterlife, shaped less by individual intention and more by patterns of visibility, repetition, and recognition.

¹⁶ Aby Warburg, *The Renewal of Pagan Antiquity: Contributions to the Cultural History of the European Renaissance*, ed. and trans. David Britt (Los Angeles: Getty Research Institute, 1999)

2.3 Loops, Repetition, and Cybernetic Shame



Figure 17. Still from a TikTok video by @pinkydollreal, Showing “Pinky Doll” performing a cowboy gesture (TikTok, 2025).

Figure 18. Still from a TikTok video by @pinkydollreal, Showing Pinky Doll performing a hand to mouth gesture (TikTok, 2025).

The character faces the camera, slightly bouncing while rotating around its vertical axis. The arms do not rest naturally along the body but are held slightly away from the torso, as if permanently prepared for movement. The gaze seems aimed toward the viewer, yet never quite meets it. At times, an arm is waved above the head in a cowboy-like gesture; at others, fingers decorated with long acrylic nails poke at the air. A flicking motion follows, after which the hand briefly covers the mouth in what appears to be a gesture of shyness. Fragmented words are uttered: “Ice cream.” “Yes yes yes.” “Gang gang.”.

The general body language is smooth and continuous. Movements are swift but not abrupt, fluid without being expressive. The character’s appearance is unmistakably human, and yet the uncanny sensation produced by a Pinky Doll performance seems to arise from a disjunction between familiar gestures and their orchestration into an eerie, dissonant whole. The intention suggested by the cowboy arm does not align with the tense and immobile spine. Facial expressions fail to respond to the movements of the hands. Gestures and

words repeat on a loop, allowing the viewer to encounter them multiple times without the temporal space necessary to process or integrate them.

How can we account for the morbid fascination produced by this double appropriation: a human performing as an algorithm-generated character designed to perform like a human? Are we amused by technology's failure to produce a convincing NPC, or unsettled by the realization that these gestures have been extracted from human behavior and mirrored back to us so thoroughly that their origin becomes indistinguishable? The uncanny, as Freud describes it, emerges when something familiar becomes strangely threatening, often through compulsive repetition¹⁷. Repetition, in this sense, exposes a loss of agency and suggests the presence of forces operating beyond conscious intention.

This logic resonates with the defining condition of the NPC, whose unease lies precisely in the absence of agency. The morbidity of the performance stems from the sense that gestures are no longer motivated by desire or will, but by automation—as if the body were a puppet animated by a non-human system. The looped gestures do not produce meaning or progression; instead, they interrupt it, suspending the character in a state of perpetual almost-action.

Eve Kosofsky Sedgwick and Adam Frank, building on Silvan Tomkins, conceptualize shame not merely as an emotion but as a mode of interruption¹⁸. Shame arises when a body's expressive energy is halted, rendering the self simultaneously hyper-visible and hyper-conscious. It is fundamentally relational: shame exists in the feedback loop between a perceiving self and an imagined or actual audience. In the context of short-form video, this dynamic is mirrored and amplified. The Pinky Doll NPC enacts continuous loops of gestures, postures, and fragmented speech that simultaneously signal intention and deny it. Movements appear deliberate yet are automated, stripped of conventional purpose or consequence. Viewers experience this as a form of cybernetic shame: the body's energy is legible but continually interrupted by its artificiality, producing fascination, discomfort, and suspended agency.

This gap between outward gesture and underlying intention can be further illuminated by Lisa Feldman Barrett's constructionist theory of emotion. In her Pufendorf Lecture, Barrett argues that emotions are not located in the face or the body as stable, readable signals¹⁹. Physical movements, she insists, have no intrinsic emotional meaning; they become meaningful only through contextual interpretation. A smile may signify joy, irony, or grief; stillness may signal calm or contained rage. Emotion, in this sense, is culturally constructed and interpreted. This claim becomes especially potent at a time when AI-powered facial detection systems deployed across advertising, education, surveillance, and interface design increasingly operate on the opposite premise: that emotions can be inferred directly from facial expressions.

What emerges across these performances is not merely a strange aesthetic or a novel genre, but a structural condition: gesture caught in a loop. Once stripped of intention, context, and consequence, bodily movements become optimizable units—repeatable,

¹⁷ Sigmund Freud, *The Uncanny*, trans. David McLintock (London: Penguin, 2003).

¹⁸ Eve Kosofsky Sedgwick and Adam Frank, "Shame in the Cybernetic Fold," in *Touching Feeling: Affect, Pedagogy, Performativity* (Durham, NC: Duke University Press, 2003)

¹⁹ Lisa Feldman Barrett, "How emotions are made" Pufendorf Lecture, Lund University, May 2024.

iterable, and endlessly recirculated. In this environment, expression no longer unfolds toward meaning or resolution, but feeds back into the system that solicits it. The discomfort provoked by the NPC, the Pinky Doll, or the looping performer does not stem from excess emotion, but from its suspension within a closed circuit. To understand this condition more fully, it is necessary to shift from gesture as expression to repetition as structure—from isolated performances to the feedback loops that sustain them.

2.4 From Repeated Gesture to Atmosphere

When gestures no longer belong to characters but circulate as repeatable units—shrugs, blinks, pauses, routines—they cease to function narratively. Accumulated across feeds, they begin to function atmospherically. What emerges is not a story world but a vibe environment: an affective landscape that envelops both character and viewer. The main character and the NPC are nothing without a vibe to move through—it gives their gestures coherence and legibility. Yet this vibe does not belong to the character, nor fully to the world they inhabit. It is the condition of both their existence.

This becomes particularly visible in the 2024 case in which TikTok influencer Sydney Sloneker accused another influencer of stealing her “vibe”. Both produced content aligned with the so-called *clean girl* aesthetic: dewy skin, minimal makeup, neutral tones, and ritualized self-care framed as effortless. After spending time together, Sloneker noticed that the other influencer had begun reproducing not only her visual style but her locations, compositions, poses and hair color. Although the case was dropped, it raised a provocative question: can a vibe be owned?

The intensity of this dispute suggests that a vibe is not simply a personal style but a form of collective attunement. What was perceived as theft was not the copying of individual actions, things, or symbols, but the appropriation of an entire affective environment—one that gains value precisely because it feels shared, recognizable, and repeatable.

Could it be that the desire to “own” a vibe may thus express less a claim to originality than a longing for participation in a collective, a shared reality? To understand this shift, it is useful to turn to Atmosphere theory, as developed by Gernot Böhme and expanded within affect theory²⁰. Atmospheres are not objects of perception but shared conditions that arise between perceiver and perceived. They act upon us, modulating feelings and attention before interpretation takes place. In this sense, atmospheres are inherently collective: they exceed individual bodies while remaining deeply felt.

Friedlind Riedel describes atmospheres as forces that temporarily unify diverse bodies into a seamless affective field²¹. Yet this coherence is not neutral. Atmospheres can simulate unity, erase difference, and normalize behavior by imposing a shared tone onto otherwise unrelated elements. What appears as harmony may in fact be a form of soft

²⁰ Gernot Böhme, *The Aesthetics of Atmospheres*, ed. Jean-Paul Thibaud (New York: Routledge, 2016).

²¹ Friedlind Riedel, “Atmosphere”, in *Affective Societies: Key Concepts*, ed. Jan Slaby and Christian von Scheve (London: Routledge, 2019)

control. This takes all its significance in relation to the emerging popular concept of “vibe politics” (the idea that in contemporary politics, aesthetic coherence often outweighs ideological clarity).

Thinking of Gesture as the site where atmospheric power is activated in User-generated short-form video, it is interesting to recall that movement is at the core of Aby Warburg’s concept of the Pathosformel. A pathos formula is not a static image or symbol, but a relational formula of affect—a repeatable configuration of movement and intensity that survives across time. It is not a quantity, but a relation: a way in which affect can be stored, translated, and reactivated in different historical contexts.

Crucially, Warburg locates affective intensity not in facial expression or narrative action, but in movement itself. In his analysis of Botticelli, he focuses on what he calls *Bewegtes Beiwerk*—movements expressed by flowing hair and fluttering garments²². These elements add little to narrative meaning, yet they intensify the image by creating an illusion of motion and liveliness. This suggests that affect resides in these movements, independent of narratives or character psychology.



Fig. 19. Signaling (2025), video installation, HD video still showing sleeping figures in public space, 16:9, 1 min loop. Image courtesy of the artist.

²² Aby Warburg, *Sandro Botticelli's Birth of Venus and Spring* (1893), trans. David Britt, in *The Renewal of Pagan Antiquity: Contributions to the Cultural History of the European Renaissance*, intro. Kurt W. Forster (Los Angeles: Getty Research Institute, 1999)



Fig. 20. *Signaling* (2025), video installation, HD video still showing a sleeping figure in public space , 16:9, 1 min loop. Image courtesy of the artist.

In *Signaling*, this atmospheric logic is not illustrated but enacted. Gestures that initially function as narrative promises—the knife lick, the fearful gaze, the crouching body—lose their directional force and begin to operate as circulating affective units. As viewers acclimate to the rhythm of return, suspense thickens without release, giving way to atmosphere. What remains is a vibe sustained by cinematic intensity rather than story: an immersive field in which characters persist as carriers of affect, moving within an environment that feels charged, familiar, and compelling, yet fundamentally undecidable. *Signaling* thus stages atmosphere not as background or mood, but as the primary mode through which gesture survives after narrative has stalled.

Synthesis

This thesis set out to examine how User-generated short-form video operates as a dominant contemporary media form by privileging affect, gesture, and atmosphere over narrative development, dialogue, or resolution. Through an analysis of archetypal figures such as the Main Character and the NPC, and through concepts drawn from affect theory, atmosphere theory, and art history, I have argued that User-generated short-form video produces a post-narrative sensibility in which legibility is no longer primarily organized through narrative causality, but through repetition, recognizability, and emotional modulation. In this context, gestures detach from individual intention and circulate as formulas, while vibes emerge as collective affective environments that shape perception in advance of interpretation.

By mobilizing Aby Warburg's notion of *Pathosformel*, this thesis reframes this media form not as a rupture from historical visual culture, but as a contemporary intensification of inherited affective logics. Just as pathos formulas migrate across time through repeated gestures and moving accessories, User-generated short-form video recycles bodily expressions, poses, and moods that retain affective force even as their narrative anchoring dissolves. The loop, central to both algorithmic platforms and NPC performance, functions as a key mechanism in this process: it suspends progression, erodes causality, and transforms time into a continuous present.

Eve Kosofsky Sedgwick's understanding of shame as interruption further clarifies how these media objects affect viewers. Rather than producing catharsis or identification, short-form video frequently generates moments of affective rupture—pauses, glitches, repetitions—that disrupt feedback and expose the viewer's own position within the circuit of attention. These interruptions do not resolve into critique or transformation; instead, they sustain engagement through disorientation, fascination, or numbness. Transgression, as shown through NPC and Main Character performances, exists largely without consequence, functioning as spectacle rather than risk.

Understanding this shift requires taking seriously forms of affective experience that do not stabilize into interpretation. Interruption, repetition, and exposure generate modes of attention that are neither opaque nor transparent, but provisional—continually recalibrated through affect rather than secured through narrative resolution.

Bibliography

- Barrett, Lisa Feldman. "How Emotions Are Made." Pufendorf Lecture, Lund University, May 2024.
- Bishop, Claire. *Disordered Attention: How We Look at Art and Performance Today*. London: Verso, 2024.
- Böhme, Gernot. *The Aesthetics of Atmospheres*. Edited by Jean-Paul Thibaud. New York: Routledge, 2016.
- Estériola, Ana. *Reproduced*. Vidéo installation, 2017.
- Freeman, Elizabeth. *Time Binds: Queer Temporalities, Queer Histories*. Durham, NC: Duke University Press, 2010.
- Freud, Sigmund. *The Uncanny*. Translated by David McLintock. London: Penguin Books, 2003.
- Goffman, Erving. *The Presentation of Self in Everyday Life*. London: Penguin Books, 1990.
- Gunawardena, Charlotte N., and Frank J. Zittle. "Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment." *American Journal of Distance Education* 11, (1997).
- Kermode, Frank. *The Sense of an Ending: Studies in the Theory of Fiction*. Oxford: Oxford University Press, 1967.
- Lambert-Beatty, Carrie. "Make-Believe: Parafiction and Plausibility." *October* 129 (2009)
- Mayer, Richard E. *Multimedia Learning*. 2nd ed. Cambridge: Cambridge University Press, 2009.
- Moreno, Roxana, and Richard E. Mayer. "Engaging Students in Active Learning: The Case for Personalized Multimedia Messages." *Journal of Educational Psychology* 92, no. 4 (2000).
- Riedel, Friedlind. "Atmosphere." In *Affective Societies: Key Concepts*, edited by Jan Slaby and Christian von Scheve, 85–95. London: Routledge, 2019.
- Sedgwick, Eve Kosofsky. *Touching Feeling: Affect, Pedagogy, Performativity*. Durham, NC: Duke University Press, 2003.
- Sedgwick, Eve Kosofsky, and Adam Frank. "Shame in the Cybernetic Fold." In *Touching Feeling: Affect, Pedagogy, Performativity*, 93–122. Durham, NC: Duke University Press, 2003.
- Steyerl, Hito. "Is a Museum a Factory?" In *The Wretched of the Screen*, 69–75. Berlin: Sternberg Press, 2012.

Su, Chunmei. *Douyin, TikTok and China's Online Screen Industry: The Rise of Short Video Platforms*. New York: Routledge, 2022.

Tomkins, Silvan S. *Affect Imagery Consciousness*. Vol. 3, *The Negative Affects: Anger and Fear*. New York: Springer, 1991.

Warburg, Aby. *The Renewal of Pagan Antiquity: Contributions to the Cultural History of the European Renaissance*. Edited and translated by David Britt. Los Angeles: Getty Research Institute, 1999.

———. "Sandro Botticelli's *Birth of Venus* and *Spring* (1893)." Translated by David Britt. In *The Renewal of Pagan Antiquity: Contributions to the Cultural History of the European Renaissance*, introduction by Kurt W. Forster, 89–126. Los Angeles: Getty Research Institute, 1999.

Wark, McKenzie. *Gamer Theory*. Cambridge, MA: Harvard University Press, 2007.