Procedural Generation of Attention: Gaze Trajectories and the Grammar of Cinematic Space

Flyxion

Independent Researcher

June 2026

Abstract

We propose a formal theory of *admissibility-guided traversal* and use cinema as its primary motivating example. The central claim is that designed systems generate experience by shaping admissibility structures $(\mathcal{X}, \mathcal{A})$ that constrain and guide trajectories through possibility spaces. A designer constructs $(\mathcal{X}, \mathcal{A})$; an observer generates a trajectory γ satisfying $\gamma(t+1) \in \mathcal{A}(\gamma(t))$; experience emerges from the properties of γ rather than from the terminal state $\gamma(T)$ alone. This structure — which we call a *traversal system* — is instantiated by cinema, game design, musical composition, and mathematical proofs, among other systems.

In cinema, the state space is the visual fixation field t, and $\mathcal{A}$ is determined by salience gradients, saccadic transport cost, and editing structure. We show that the gaze density field evolves according to a Fokker–Planck advection-diffusion equation driven by the salience function, that camera movement induces a vector-field coupling whose stability is governed by the smooth pursuit threshold, and that classical editing conventions — the 180-degree rule, eyeline matching, the 30-degree rule — emerge as approximate solutions to a re-acquisition minimization over gaze anchor density, derivable from perceptual principles without reference to narrative convention.

We then establish the *Traversal Reward Decomposition*: $R(\gamma) = R(\gamma(T)) + R_T(\gamma)$, where R_T is a traversal reward that is non-zero for non-degenerate admissibility structures and independent of informational novelty. This explains why replay, rewatching, rereading, rehearsal, and ritual are not anomalous. Wim Wenders' *Perfect Days* is examined as a natural experiment in which $R_T \gg R(\gamma(T))$, isolating traversal reward from terminal reward. The *Traversal Equivalence* theorem shows that cinema, games, music, and proofs are structurally identical at the level of $(\mathcal{X}, \mathcal{A}, \gamma)$; the film–game correspondence table follows as a corollary rather than an analogy. The *Designer–Observer Separation* theorem formalizes the bilateral structure of designed experience: the designer controls $\mathcal{A}$; the observer generates γ; neither fully determines experience unilaterally. The framework generates empirical predictions testable by eye-tracking and psychophysiological methods and points toward a general account of admissibility-structured cognition that extends from perceptual attention through play, ritual, and education.

Contents

1 Introduction

Film theory traditionally begins with narrative. Attention research begins with fixation. Game studies begins with interaction. This paper begins with traversal.

The three fields have each generated substantial knowledge about designed experience. Film theory has illuminated how stories are structured, how meaning is constructed across shots, and how editing produces coherence from discontinuity. Attention research has documented how the eye moves through visual fields, what salience factors drive fixation, and how perceptual systems extract information from scenes. Game studies has developed accounts of interactivity, player agency, and the design of navigable spaces. What none of these traditions has done is identify the common substrate that underlies all three.

That substrate is traversal. We propose that cinema, game design, music, and other forms of designed experience are all instances of the same formal structure: a designer constructs an admissibility structure $(\mathcal{X}, \mathcal{A})$ over a state space, and an observer generates a trajectory γ through that structure satisfying $\gamma(t+1) \in \mathcal{A}(\gamma(t))$. Experience emerges from properties of the trajectory rather than from the terminal state alone. The designer does not control the observer's path. The designer controls what paths are possible.

$$(\mathcal{X}, \mathcal{A}) \;\longrightarrow\; \gamma \;\longrightarrow\; R(\gamma) \tag{1}$$

Diagram (1) is the organizing structure of the paper. The left arrow is the designer's contribution: specifying a state space and a transition function. The middle arrow is the observer's contribution: generating a trajectory by local admissible steps. The right arrow is experience: the reward, affect, and meaning that emerge from properties of the trajectory rather than from its endpoint.

We begin with gaze because it is measurable. Eye-tracking technology makes gaze trajectories directly observable, and the extensive literature on visual attention provides a solid empirical foundation for the spatial claims of Sections 2 and 3. But gaze is the entry point, not the subject. The framework developed here applies equally to locomotion through a game level, expectation through a musical phrase, and inference through a mathematical proof. The proof case is the strongest demonstration that the theory is genuinely about traversal: when

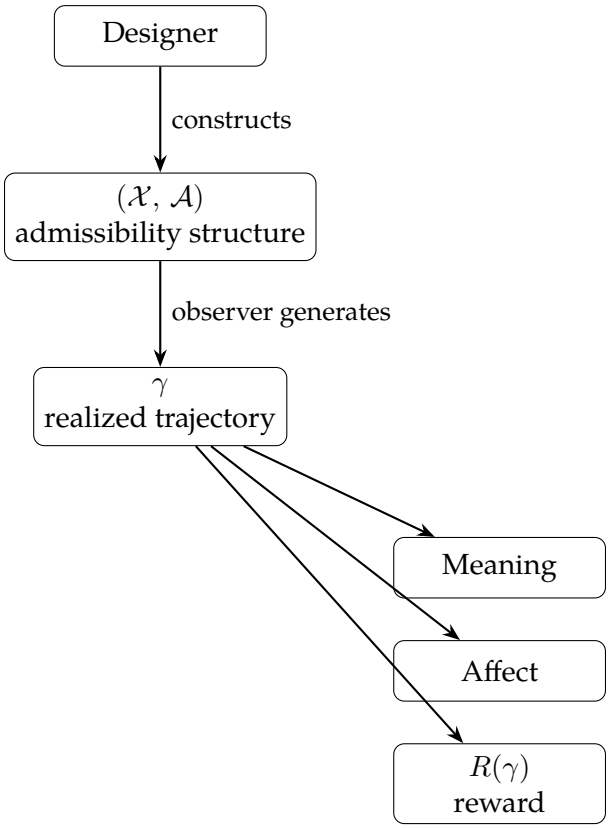

Figure 1: The organizing structure of the traversal framework. The designer constructs $(\mathcal{X}, \mathcal{A})$; the observer generates γ by local admissible steps; meaning, affect, and reward emerge as properties of γ rather than of the terminal state $\gamma(T)$.

the same formal structure that describes gaze control in cinema also describes the experience of reading a known proof, the framework has reached something more general than vision.

The paper proceeds as follows. Section 2 introduces the screen as a weighted fixation field and describes cinematic devices as geometric operations on that field. Section 3 develops the theory of gaze trajectories, formalizes the cut as a discontinuity event, derives editing grammar from a re-acquisition minimization, and states the procedural generation claim precisely. Section 4 distinguishes spatial from temporal guidance and shows that camera movement, editing rhythm, object motion, and musical phrasing are all operations on the same temporal geometry. Section 5 shows how the spatial and temporal structures of the preceding sections give rise to recurring affective operations, without proposing a general theory of emotion. Section 6 states the framework's em-

pirical predictions. Section 6 steps back to show that film, games, music, and proofs are all traversal systems in the sense of the formal definition, locates prior research traditions within the framework, and argues that traversal is intrinsically rewarding independent of informational novelty. Section 7 returns to the traversal claim as a demonstrated result.

2 The Screen as Spatial Field

2.1 The Frame as a Weighted Fixation Field

A frame contains objects. This is how films are ordinarily described: a room, two characters, a gun on the table. But this description operates at the wrong level for the present analysis. Before the viewer identifies any object, before narrative content is assigned to anything on screen, the frame has already done something more primitive: it has defined a distribution over possible fixation points. The filmmaker's first task is not to tell a story but to shape the topology of the fixation field.

Definition 2.1 (Gaze Field)**.** At time t, the *gaze field* is a pair (t, w_t) where $t = \{g_1, g_2, \ldots, g_n\}$ is the set of candidate fixation points in the current frame and $w_t : t \to [0, 1]$ is a salience function assigning to each point a weight proportional to the probability that an unconstrained viewer will fixate it. Salience is determined jointly by motion, luminance contrast, edge density, face presence, implied gaze vectors, text, and color singularity.

The salience function w_t is not uniform. In any given frame, some regions attract fixation far more strongly than others. A face in a scene of ambient background will draw the eye with near-universal consistency across viewers. A moving object in an otherwise static frame will capture attention before its identity is recognized. These are not aesthetic preferences but perceptual facts about the human visual system, extensively documented in eye-tracking research [1, 2].

What follows from this is that every compositional decision — where to place actors, where to direct light, what aperture to use, how to set the camera height

— is a decision about (t, w_t). The composition is not primarily a picture. It is a distribution. Two frames containing identical narrative content but different compositions produce different distributions, direct the eye differently, and produce different experiences even when a viewer could describe both in the same words.

We can quantify the degree of attention control a given frame achieves by measuring the entropy of its salience distribution.

Definition 2.2 (Salience Entropy)**.** The *salience entropy* of a frame at time t is

$$S(t) = -\sum_{g_i \in t} w_t(g_i) \log w_t(g_i). \tag{2}$$

$S(t)$ is maximized when w_t is uniform over t and minimized when w_t concentrates all mass on a single fixation point.

Proposition 2.3 (Salience Concentration)**.** *Rack focus, localized illumination, and foreground isolation each decrease $S(t)$. Consequently these operations reduce the effective number of admissible fixation targets.*

Proof. Each operation transfers probability mass from a broad distribution over t to a concentrated distribution over a proper subset $'_t \subsetneq t$. Since entropy is maximized by the uniform distribution and strictly decreases under mass concentration, $S('_t) < S(t)$. The effective fixation set — the number of candidates with non-negligible salience weight — contracts accordingly. □

Proposition 2.3 makes precise what it means to say that a rack focus or key light "controls" viewer attention. Control is not binary. It is a continuous reduction in $S(t)$, measurable in principle from the salience distribution induced by any given composition. A perfectly controlled frame has $S(t) \approx 0$; a completely uncontrolled frame has $S(t) = \log |t|$. Most film compositions operate in the middle of this range, with deliberate moments of high control (a face in a spotlight) and deliberate moments of low control (an establishing wide shot inviting exploration).

The salience function is not merely a static weighting. In any continuous shot it changes over time as objects move, light shifts, and the camera reframes. To capture this dynamics we model gaze density as a field evolving under salience.

Definition 2.4 (Gaze Density Field)**.** The *gaze density field* $\rho(g, t)$ is a probability density over t representing the distribution of fixations across a population of viewers at time t, normalized so that $\int_t \rho(g, t)\, dg = 1$.

Theorem 2.5 (Salience Flow)**.** *Let $w_t : t \to [0, 1]$ be continuously differentiable in both g and t. Then the gaze density field $\rho(g, t)$ evolves according to the advection-diffusion equation*

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \big(\rho\, \nabla w_t\big) + D\, \Delta \rho, \tag{3}$$

where $D \geq 0$ is an exploration parameter governing the rate of diffusion of gaze away from salience peaks. Consequently, regions of high salience act as attractors of fixation density: in the limit $D \to 0$, ρ concentrates on the maxima of w_t.

Proof. Model each viewer's fixation as a particle moving in the salience landscape under gradient ascent perturbed by Brownian noise: $d\gamma = \nabla w_t\, dt + \sqrt{2D}\, dW_t$. The Fokker–Planck equation for the density of such particles is exactly (3). In the zero-noise limit $D \to 0$ the diffusion term vanishes and the flow is pure gradient ascent, concentrating mass on salience maxima. □

Theorem 2.5 connects the static analysis of Section 2.1 to the dynamic trajectory analysis of Section 3. The filmmaker who reshapes w_t over time — through movement within the frame, lighting changes, or camera motion — is modifying the vector field ∇w_t that drives ρ. This is not merely influencing where viewers look. It is specifying a dynamical system whose attractors are the filmmaker's intended fixation targets.

2.2 Cinematic Devices as Geometric Operations

The tools available to a cinematographer can be understood as a set of operations on the gaze field (t, w_t). Describing them in these terms clarifies what they do at the perceptual level, prior to any interpretation of their narrative function.

Rack focus. A rack focus shifts the lens focal plane from one depth to another within the frame, throwing the previously sharp region into blur and bringing the newly focused region into clarity. In gaze field terms, this is a salience annihilation operation:

$$w_t(g_i) \to 0 \quad \text{for all } g_i \text{ outside the new focal plane} \tag{4}$$

The number of candidate fixation points in t does not change. The number of high-salience fixation points collapses to the focused region. The filmmaker does not tell the viewer where to look. The filmmaker destroys the salience of every alternative.

Lighting. A lighting setup shapes the salience field across the entire frame by controlling the distribution of luminance contrast. Hard directional light creates steep salience gradients, concentrating the distribution around bright regions and deep shadows. Flat ambient light produces a shallow gradient and a diffuse distribution. When a key light isolates a face in an otherwise dark frame, it performs salience concentration by luminance means rather than optical means. The effect on w_t is structurally identical to a rack focus: the viewer's eye is given little alternative to the lit region.

Blocking. The placement of actors within the frame distributes high-salience objects — faces, bodies in motion, implied gaze vectors — across t. A scene in which two actors stand at opposite edges of a wide frame creates a split distribution with two competing attractors and a low-salience region between them. A scene in which one actor stands close to camera while another recedes into background creates a depth hierarchy in w_t. The director blocking a scene is engineering this distribution: determining which regions of the frame will compete for the viewer's attention and in what order.

Framing and focal length. The choice of lens and camera distance determines how much spatial world is projected onto the frame, how much background competes with foreground for salience, and how much of t is occupied by the primary subject. A long lens compresses depth, reduces background detail, and concentrates w_t on the subject. A wide lens expands the field, increases the number of candidate fixation points, and widens the distribution.

2.3 The Game Level as Analogous Spatial Field

A game level is a navigable volume rather than a projected frame, but it operates on the player's gaze field by identical means. The level designer controls a spatial environment from which the player's visual attention must be directed toward intended targets, guided through intended paths, and managed across an extended traversal.

The toolkit is structurally the same. Lighting in a game environment concentrates salience exactly as lighting on a film set does: a bright opening at the end of a dark corridor draws the player's eye and locomotion toward it precisely as a key light draws the viewer's gaze to a face. Architectural funneling — the use of converging walls, narrowing passages, and forced perspectives — constrains t spatially, reducing candidate fixation directions and biasing the player toward the intended path. A highlighted interactable object, marked by a glow, a distinctive color, or an animation, is a salience concentration operation identical to a rack focus: the designer destroys the relative salience of the surrounding environment to force attention to the target.

The differences between a film frame and a game level are real but secondary from this perspective. A frame is a two-dimensional projection; a level is a three-dimensional volume. A film viewer cannot redirect the camera; a player can redirect their gaze within the available volume. But in both cases the designer's task is the same: construct a salience field (t, w_t) that channels attention toward intended regions while leaving the observer the subjective experience of choosing where to look. Effective spatial design in both film and games produces the experience of free attention while delivering constrained attention. The observer feels they are looking where they choose. They are looking where the designer arranged.

3 Trajectory Generation

3.1 Gaze Trajectories

The gaze field (t, w_t) defined in Section 2.1 specifies where the eye *could* go at any moment. It does not specify where the eye *will* go. Between these two ques-

tions lies the theory of gaze trajectories: the study of how the perceptual system generates a path through the fixation field rather than merely responding to individual frames.

Definition 3.1 (Gaze Trajectory)**.** A *gaze trajectory* is a path $\gamma : [0, T] \to$ through fixation space, where $= \bigcup_t t$ is the total fixation space across all frames. At each moment t, the trajectory specifies the current fixation $\gamma(t) \in t$. The set of locally admissible continuations from $\gamma(t)$ is denoted $\Gamma(\gamma(t))$.

The critical property of gaze trajectories is that they are not arbitrary paths through. The eye does not sample the fixation field randomly. At each moment, the next fixation is generated from the current fixation by a process that is simultaneously bottom-up (driven by the salience field) and top-down (constrained by task, history, and expectation). This dual determination is the source of the filmmaker's leverage: by controlling w_t, the filmmaker shapes the bottom-up component of trajectory generation even when the top-down component varies across viewers.

Proposition 3.2 (Local Trajectory Constraint)**.** *Let $\gamma(t)$ denote the current fixation. Then the probability distribution over the next fixation $\gamma(t+1)$ is jointly determined by:*

(i) local salience gradients ∇w_t in the neighborhood of $\gamma(t)$,

(ii) saccadic transport cost $d(\gamma(t), g_i)$ for each candidate $g_i \in t$,

(iii) task constraints specifying which regions are task-relevant at time t, and

(iv) prior fixation history $\gamma(0), \ldots, \gamma(t-1)$, encoding inhibition of return and scene memory.

Consequently, gaze trajectories are locally constrained traversals of, not arbitrary paths. The admissibility structure $\mathcal{A}$ governing trajectory generation is determined by factors (i)–(iv) jointly.

Proposition 3.2 has an important corollary for the theory of cinematic control. The filmmaker directly controls factor (i) through cinematography and composition. Factors (ii) is a fixed property of the oculomotor system. Factor

(iii) is shaped partly by the filmmaker through narrative cues that establish task relevance. Factor (iv) is a function of the film's prior shots and therefore also under the filmmaker's indirect control. This means the filmmaker has leverage over three of the four determinants of the next fixation. The impression that the viewer is freely directing their own gaze is accurate at the phenomenological level and misleading at the mechanistic level.

Two consequences follow for the structure of the paper. First, the transition from gaze fields to gaze trajectories is not a change of subject but a deepening of the same analysis: the field specifies the distribution at a moment, the trajectory specifies how that distribution is traversed over time. Second, the four-factor structure of Proposition 3.2 implies that trajectory generation is a genuinely local process — each fixation is determined by the current state and immediate neighborhood, not by global knowledge of the film — which is precisely the condition under which procedural generation in the sense of Section 3.5 is the right description of what is happening.

The local character of trajectory generation also implies a structural fact about gaze paths that motivates the special status of cuts.

Proposition 3.3 (Piecewise Continuity)**.** *For sufficiently small temporal intervals Δt between saccadic events,*

$$d\big(\gamma(t+\Delta t),\, \gamma(t)\big) < \varepsilon \tag{5}$$

for some ε bounded by oculomotor constraints. Consequently, gaze trajectories are piecewise continuous paths through $\mathcal{V}$, with discontinuities occurring only at saccadic events.

Proof. Between saccades, the eye is in fixation: its angular position changes by at most the drift rate of the oculomotor system, which is small relative to the scale of the fixation field. The bound ε is therefore determined by physiology rather than by the film. Discontinuities in γ occur only when a saccade is initiated, which is a discrete event with non-zero inter-saccade intervals. □

Proposition 3.3 is the formal grounding for why cuts are special. Within a continuous shot, gaze trajectories are continuous: the eye traverses t smoothly, guided by the salience field and the camera motion field V_t. A cut introduces a discontinuity in t itself, not merely in γ: the frame changes instantaneously.

The perceptual system must then generate a new fixation from a standing start, which is the re-acquisition problem formalized in Section 3.3. Camera movements that are continuous in t are therefore qualitatively different from cuts, not merely faster or slower versions of the same event.

The local selection of the next fixation point can be stated as an explicit optimization.

Proposition 3.4 (Least-Effort Saccade Principle)**.** *The expected next fixation satisfies*

$$\gamma(t+1) = \arg\max_{g_i \in t} \Big(w_t(g_i) - \lambda \, d\big(\gamma(t), g_i\big) \Big), \tag{6}$$

where $\lambda > 0$ is a cost coefficient weighting saccadic transport distance against informational gain. Fixation selection therefore balances the pull of salience against the cost of movement.

Proof sketch. Equation (6) follows from treating the oculomotor system as approximately optimizing an immediate expected utility: the benefit of fixating g_i is proportional to the information potentially gained, which is captured by $w_t(g_i)$, and the cost is the metabolic and temporal cost of the saccade, which is proportional to angular distance $d(\gamma(t), g_i)$. Under a linear utility model the optimizer takes the stated form. This is consistent with empirical models of saccadic selection [2] and with the inhibition-of-return phenomenon, which reduces $w_t(g_i)$ for recently visited locations. □

Proposition 3.4 has a direct consequence for cinematography. The filmmaker who places a high-salience target near the current expected fixation position imposes lower transport cost and therefore higher probability of rapid acquisition. Placing a high-salience target at the far edge of the frame simultaneously raises the salience term and the cost term in (6). Whether the target is acquired depends on the ratio $w_t(g_i)/(\lambda \, d(\gamma(t), g_i))$. The filmmaker managing this ratio across a scene is implicitly solving the same optimization that the viewer's oculomotor system is solving locally.

3.2 Camera Movement as Trajectory Imposition

Camera movement acts on the gaze trajectory by introducing a systematic, time-varying displacement of the entire fixation field. Where the operations of Section 2.2 modified the salience distribution within a static frame, camera movement modifies the frame itself: it translates, rotates, scales, or perturbs t as a whole, imposing a motion on every candidate fixation point simultaneously.

We formalize this as a vector field on the gaze space.

Definition 3.5 (Camera Motion Field)**.** At time t, a camera movement induces a *camera motion field* $V_t : t \to Tt$, assigning to each candidate fixation point $g_i \in t$ a displacement vector $V_t(g_i)$ in the tangent space of t. The induced trajectory displacement is $\gamma(t + dt) = \gamma(t) + V_t(\gamma(t)) \cdot dt$ in the limit of continuous camera motion.

Each camera move corresponds to a characteristic class of vector field on t.

Pan. A horizontal pan translates the entire gaze field laterally at a uniform rate. The field V_t is approximately constant across t: every fixation point is displaced by the same vector. The pan carries the eye along a constrained horizontal path regardless of which fixation point within the field it currently occupies. This is trajectory imposition in the clearest sense: the viewer can resist the pan by fixating against its direction, but the salience field is also moving, so resistance requires active effort and produces the experience of looking away from what the film wants to show.

Dolly and tracking. A dolly or tracking shot moves the camera through physical space, producing a parallax displacement that varies with depth: nearby objects appear to move rapidly across the field while distant objects move slowly. The field V_t is therefore non-uniform, with magnitude increasing as a function of proximity. The viewer experiences this as locomotion: the visual system interprets depth-dependent parallax as self-motion through a stable environment. This is why the dolly and the zoom, which produce superficially similar screen effects, feel phenomenologically different. The dolly provides the full parallax signature of locomotion and triggers the proprioceptive correlates of movement. The zoom does not.

Zoom. A zoom changes focal length without moving the camera, scaling the entire image toward or away from its center. The induced field V_t is a radial expansion or contraction centered on the frame center: $V_t(g_i) \propto (g_i - c)$ where c is the center point. The zoom produces the visual appearance of approach or retreat without the parallax signature of actual locomotion. The perceptual system receives conflicting information: the image says approach, the vestibular system says stationary. The characteristic affective quality of the zoom — unease, dream-like dislocation, the sense of something wrong with the movement — follows directly from this conflict between the optical and proprioceptive channels. The dolly and zoom are not stylistic variants. They are geometrically distinct operations producing different sensorimotor signatures.

Handheld. A handheld camera introduces stochastic perturbation into V_t: the field is no longer smooth but noisy, with high-frequency random displacements overlaid on any intentional camera movement. In trajectory terms, this widens the locally admissible fixation set at each moment: the viewer cannot predict exactly where a given region of interest will be in the next instant. The uncertainty is not merely informational. It is perceptual: the visual system must continuously re-acquire its target as the field perturbs around it. The physical correlate of handheld imagery — bodily instability, urgency, documentary authenticity — arises because the perturbation pattern matches the signature of a moving, unstable observer rather than a fixed camera platform.

The camera motion field V_t is the filmmaker's most direct instrument for trajectory imposition because it acts on the gaze trajectory independently of where within the frame the viewer happens to be looking. Salience shapes the distribution over t; V_t moves t itself. The two instruments operate at different levels of the trajectory generation process and can be combined or opposed to produce complex effects: a pan that moves the salience field in one direction while a bright object appears at the frame edge in the opposite direction creates a competition between V_t and w_t, and the tension of that competition is experienced directly by the viewer as a pull between two attentional demands.

The relationship between V_t and the resulting gaze trajectory can be stated as a formal alignment claim.

Proposition 3.6 (Trajectory Alignment)**.** *Let $V_t : t \to Tt$ be a coherent motion field induced by camera movement. Then expected gaze velocity aligns with V_t up to bounded error:*

$$\mathbb{E}\big[\dot{\gamma}(t)\big] = \alpha V_t(\gamma(t)) + \epsilon_t \tag{7}$$

where $\alpha > 0$ is a coupling coefficient determined by the relative strength of camera motion against competing salience signals, and ϵ_t is a zero-mean error term bounded by oculomotor noise and task-driven deviations.

Proof sketch. A coherent camera motion field displaces every candidate fixation point in t by $V_t(g_i) \cdot dt$ in a small interval dt. If the viewer is currently fixating $\gamma(t)$, the target of the fixation moves to $\gamma(t) + V_t(\gamma(t)) \cdot dt$. The smooth pursuit system of the oculomotor apparatus tracks moving targets at velocities up to approximately $30^\circ\,\mathrm{s}^{-1}$; within this range, gaze velocity matches target velocity with high fidelity. The coupling coefficient α falls below unity when camera speed exceeds the smooth pursuit limit, inducing corrective saccades. The error term ϵ_t accumulates competing salience signals that may redirect gaze against the camera motion. □

Proposition 3.6 formally states what pans and tracking shots accomplish: they induce a coherent V_t that the oculomotor system tracks, carrying the gaze trajectory with the camera within the bounds of the smooth pursuit system. The dolly, pan, tilt, and tracking shot are all instances of coherent V_t fields with different geometric structures; the handheld camera is a stochastic V_t that disrupts alignment and forces increased ϵ_t.

When camera speed remains below the smooth pursuit threshold, gaze does not merely align with camera motion on average — it locks onto the moving target asymptotically.

Theorem 3.7 (Tracking Stability)**.** *Suppose $\|V_t\| < V_c$ for all t, where V_c is the smooth-pursuit velocity threshold of the oculomotor system. Let $g^*(t)$ denote the intended target position moving under V_t. Then*

$$\lim_{t\to\infty} \big\|\gamma(t) - g^*(t)\big\| = 0. \tag{8}$$

That is, sufficiently smooth camera motion induces asymptotically stable gaze locking onto the moving target.

Proof. Under the smooth pursuit system, the oculomotor error $e(t) = \gamma(t) - g^*(t)$ evolves according to a first-order error-correction loop: $\dot{e} = -ke + \eta(t)$ where $k > 0$ is the pursuit gain and $\eta(t)$ is bounded noise. When $\|V_t\| < V_c$, the pursuit system does not saturate and maintains $k > 0$. The homogeneous solution $e(t) = e(0)e^{-kt}$ decays to zero, and the forced response due to η remains bounded. Hence $e(t) \to 0$ as $t \to \infty$. □

Theorem 3.7 provides the mathematical explanation for why slow tracking shots feel effortless while fast pans feel effortful. Below V_c the smooth pursuit system maintains a stable lock and the viewer experiences the camera's motion as natural continuation. Above V_c the pursuit system saturates, corrective saccades become necessary, and the eye repeatedly loses and re-acquires the target — the perceptual correlate of a pan that feels too fast to follow.

3.3 The Cut as Discontinuity Event

Every cut is a discontinuity in the gaze trajectory. Where camera movement carries the eye along a continuous path through the fixation field, the cut resets the field instantaneously: the frame before the cut is replaced by a frame that may share none of its spatial structure. The viewer's perceptual system must locate a new fixation target from a standing start.

Formally, let $\gamma(t^-)$ denote the gaze position immediately before a cut and $\gamma(t^+)$ the gaze position immediately after. The cut introduces a discontinuity $\gamma(t^-) \neq \gamma(t^+)$ in general, and the perceptual system faces what we call the *re-acquisition problem*: given the new frame F_{t+1} and the prior fixation $\gamma(t^-)$, find the intended target in F_{t+1} as rapidly as possible.

Definition 3.8 (Re-acquisition Cost)**.** The re-acquisition cost $C(F_t, F_{t+1})$ of a cut from frame F_t to frame F_{t+1} is the time and attentional effort required for the viewer's gaze to reach the filmmaker's intended target region in F_{t+1}, given the fixation state $\gamma(t^-)$ at the moment of the cut.

Re-acquisition cost is not uniform across cuts. It depends on the relationship between the two frames. Specifically, it depends on the degree to which structures present in F_t survive into F_{t+1} in a form that can guide the eye toward its new target. We formalize this as the gaze anchor relation.

Definition 3.9 (Gaze Anchor Relation)**.** The gaze anchor relation $A_G(F_t, F_{t+1})$ is the set of perceptual structures that persist across consecutive frames F_t, F_{t+1}, grounding the inference that a common world underlies both and providing guidance for post-cut fixation. Anchor structures include: screen position, motion vector, luminance peak, face location, and implied gaze direction.

When $|A_G(F_t, F_{t+1})|$ is large, the viewer's eye has multiple guides into the new frame and re-acquisition is rapid. When $|A_G|$ is small, the eye must search the new frame with little guidance, re-acquisition is slow, and the cut may register as disorienting or jarring. A jump cut — a cut between two shots of the same subject from nearly the same angle — is disorienting not because it violates narrative expectations but because it destroys spatial anchor density while preserving just enough similarity that the viewer cannot determine whether a cut has occurred at all. An incoherent edit — a cut between entirely unrelated spaces with no shared structure — destroys anchor density entirely, producing confusion.

The relationship between anchor density and re-acquisition cost can be stated precisely.

Theorem 3.10 (Anchor Bound)**.** *There exists a monotone decreasing function* $f : \mathbb{N} \to \mathbb{R}^+$ *such that*

$$C(F_t, F_{t+1}) \leq f\big(|A_G(F_t, F_{t+1})|\big). \tag{9}$$

Consequently, increasing anchor density provides an upper bound on re-acquisition cost: more anchors guarantee faster target acquisition after a cut.

Proof. Each anchor structure $a \in A_G(F_t, F_{t+1})$ provides an independent cue that the perceptual system can use to locate the intended target in F_{t+1}. With $|A_G| = 0$, no cues are available and acquisition reduces to visual search over the entire frame, with cost bounded below by mean search time. With $|A_G| = k > 0$, at least one of k independent cues points toward the target region; the expected

time to locate the target decreases with k since the observer need only follow any one cue successfully. Define $f(k) = C_{\max} \cdot \prod_{i=1}^{k}(1 - p_i)^{-1}$ where p_i is the probability that the i-th anchor type successfully guides acquisition and $C_{\max}$ is the cost of uninstructed search; then f is monotone decreasing in k and provides the required bound. □

The Anchor Bound gives a monotone relationship between anchor count and re-acquisition cost. We can sharpen this into an information-theoretic statement by measuring the quality of anchors, not merely their number.

Definition 3.11 (Anchor Entropy)**.** The *anchor entropy* of a cut from F_t to F_{t+1} is

$$H_A(F_t, F_{t+1}) = -\sum_i p_i \log p_i, \tag{10}$$

where p_i is the probability that anchor $i \in A_G(F_t, F_{t+1})$ successfully predicts the location of the intended target in F_{t+1}. Low anchor entropy means one or a few anchors predict the target with high reliability; high anchor entropy means all anchors are uncertain.

Theorem 3.12 (Information-Theoretic Cut Cost)**.** *Re-acquisition cost satisfies*

$$C(F_t, F_{t+1}) \propto H_A(F_t, F_{t+1}) \tag{11}$$

up to perceptual constants. Cuts with low anchor entropy have low re-acquisition cost; cuts with high anchor entropy are costly regardless of the number of anchors present.

Proof. Model target acquisition after a cut as an optimal search problem. With anchor predictions $\{p_i\}$, the minimum expected search cost is achieved by testing anchors in decreasing order of p_i. The expected number of tests before success is $\sum_i i \cdot p_i(1 - p_1) \cdots (1 - p_{i-1})$, which is a monotone increasing function of H_A when the p_i are permuted toward uniformity. In the limit of uniform p_i (maximum entropy), all anchors are equally unreliable and expected search cost reaches its maximum; in the limit of a single $p_i = 1$ (zero entropy), the target is located immediately. The proportionality to H_A holds to first order in the entropy expansion of the search cost. □

Theorem 3.12 refines the Anchor Bound by distinguishing cuts that have many anchors but uncertain ones from cuts with few anchors that are individually reliable. An eyeline match is a low-entropy cut: the gaze direction anchor predicts the target location with high probability. A cut to a wide shot of an unfamiliar environment may have many candidate anchors, none individually reliable, producing high H_A and therefore high re-acquisition cost despite the anchor count being non-zero.

The grammar of continuity editing — the accumulated practical conventions by which professional editors join shots — can be understood as a set of heuristics for maintaining high anchor density across cuts. We argue that these conventions are not historically contingent stylistic norms. They are approximate solutions to a perceptual optimization problem. The closest antecedent in the existing literature is the attentional theory of cinematic continuity [21], which argues that editing conventions function primarily to guide and maintain viewer attention. The present framework formalizes that claim: anchor density is the quantity being maximized, and the editing rules emerge as solutions to the resulting optimization.

The optimization is:

$$\min_{F_{t+1}} C(F_t, F_{t+1}) \quad \text{subject to narrative constraint} \tag{12}$$

That is: among the shots that could follow the current shot given narrative requirements, choose the one that minimizes re-acquisition cost. Equivalently, maximize $|A_G(F_t, F_{t+1})|$.

The three most fundamental rules of continuity editing emerge directly from this optimization.

The 180-degree rule. In a scene involving two characters in conversation, the camera must remain on one side of an imaginary line connecting them. Crossing the line reverses the apparent screen direction of each character, destroying the motion-vector anchors established in prior shots. A character who was moving left-to-right now appears to move right-to-left without any corresponding narrative event. The 180-degree rule is the solution to: preserve motion-vector anchor structures across cuts involving character movement. It persists not because ed-

itors were taught it but because violations measurably increase re-acquisition cost and produce viewer confusion.

Eyeline matching. When a character looks off-screen in one shot, the following shot should show the object of their gaze from approximately the character's point of view. The eyeline match preserves two anchor types simultaneously: gaze direction (the character's look establishes a vector that the next shot satisfies) and, often, face location (the face appearing in shot B occupies the region that the look in shot A pointed toward). Violations produce the experience of a character looking at something other than what the edit implies they see, again because the anchor structure is broken.

The 30-degree rule. Two consecutive shots of the same subject should differ by at least 30 degrees of camera angle. Violations produce the jump cut: the subject appears to lurch forward in the frame because the screen position anchor is too similar to signal that a cut has occurred, yet different enough that spatial continuity is disrupted. The 30-degree minimum is an approximate threshold below which position anchors mislead rather than guide.

Proposition 3.13. *The three primary rules of continuity editing are solutions to the re-acquisition minimization* (12) *under the following correspondences: the 180-degree rule maximizes motion-vector anchor density; eyeline matching maximizes gaze-direction and face-position anchor density; the 30-degree rule prevents the degenerate case in which position anchor similarity is high enough to suppress cut detection but low enough to disrupt spatial continuity.*

Corollary 3.14 (Grammar Emergence)**.** *Any editing system trained solely to minimize re-acquisition cost $C(F_t, F_{t+1})$ across a corpus of cuts will converge toward the heuristics of continuity editing grammar, without access to narrative content or historically transmitted conventions.*

Proof. By Theorem 3.10, minimizing C is equivalent to maximizing $|A_G|$. By Proposition 3.13, the editing rules that maximize anchor density under the relevant geometric constraints are precisely the 180-degree rule, eyeline matching, and the 30-degree rule. An optimizer of C therefore converges to these rules as a consequence of the anchor bound alone. □

Corollary 3.14 converts the empirical claim about editing grammar into a theorem-derived consequence. Film grammar is not a historical accident encoding the aesthetic preferences of early Hollywood editors. It is the approximately optimal solution to a perceptual minimization problem. The prediction is that a computational system with no knowledge of film history, given only a model of human post-cut fixation behavior, should rediscover these conventions.

3.4 Game Locomotion as Explicit Trajectory Generation

In cinema, the viewer's trajectory through the gaze field is generated implicitly: the eye responds to salience signals and anchor structures that the filmmaker has engineered, but the eye's movement is not experienced as authored. In game design, trajectory generation is made explicit: the player moves through a navigable volume, and the experience of movement is foregrounded rather than concealed.

Despite this difference in phenomenology, the underlying structure is identical. The level designer constructs an admissibility structure $(\mathcal{X}, \mathcal{A})$ over a spatial volume: $\mathcal{X}$ is the set of positions the player can occupy, and $\mathcal{A}(x)$ specifies where the player can move from position x given the geometry of walls, floors, doors, and other constraints. The player generates a trajectory γ satisfying $\gamma(t+1) \in \mathcal{A}(\gamma(t))$. The designer does not control the player's path. The designer controls what paths are possible.

The toolkit by which level designers shape $(\mathcal{X}, \mathcal{A})$ maps directly onto the filmmaker's toolkit for shaping t. Corridors are forced perspectives: they concentrate the admissible trajectory set to a narrow band, exactly as a rack focus concentrates the gaze field. Landmarks serve as navigation markers: they are high-salience points that attract the player's trajectory, exactly as a lighting cue attracts the viewer's eye. Architectural funneling — the use of converging geometry to draw the player toward a specific location — is the spatial equivalent of the camera pan, which carries the eye along a predetermined path.

This correspondence suggests that level design and cinematography are the same discipline applied to different observer modalities. The filmmaker engineers gaze; the level designer engineers locomotion. Both are constructing

traversal systems.

3.5 Procedural Generation of Attention

The preceding sections have established that the filmmaker shapes the gaze field t, manages trajectory admissibility through camera movement, and controls reacquisition cost through editing. It might appear from this description that the filmmaker controls the viewer's attention. This appearance is mistaken, and the mistake matters.

The filmmaker does not determine where the viewer looks. The filmmaker constructs an admissibility structure $(\mathcal{X}, \mathcal{A})$ from which the viewer's trajectory emerges. The trajectory is generated by the viewer, not by the film.

The gaze field t introduced in Section 2.1 is a specific realization of this abstract structure: $t \subseteq \mathcal{X}$, where $\mathcal{X}$ is the full space of possible fixation states across the film. At each moment t, the current frame exposes a subset of $\mathcal{X}$ with a salience weighting w_t over that subset. The admissibility function $\mathcal{A}$ then specifies which transitions among fixation states are locally reachable, determined jointly by salience gradients, saccadic transport cost, and camera motion as formalized in Propositions 3.2 and 3.4. The cinematic sections of the paper are therefore a detailed instantiation of the general traversal framework in the specific state space of visual attention. The generalization to games, music, and proofs in Section 6 consists of replacing t with the appropriate domain-specific state space while leaving the structure $(\mathcal{X}, \mathcal{A}, \gamma)$ unchanged.

$$\gamma(t+1) \in \mathcal{A}(\gamma(t)) \tag{13}$$

The filmmaker specifies $\mathcal{A}$. The viewer generates γ. This is procedural generation in the strict sense of the term: a system of local rules from which global structure emerges through a generative process that the rule-specifier does not directly control.

The analogy with game design is precise here. A game designer does not control player movement. The designer constructs walls, corridors, lighting gradients, and landmark placements — the local admissibility constraints — and the player's path emerges from their interaction with those constraints. Two

players in the same level generate different trajectories. Two viewers of the same film generate different gaze paths. In both cases, the designed structure constrains but does not determine the observer's trajectory.

This framing resolves an apparent tension in film theory between the claim that films exert powerful control over viewer experience and the observation that viewers respond differently to the same film. Both claims are correct. The admissibility structure $(\mathcal{X}, \mathcal{A})$ strongly constrains the space of possible trajectories while leaving the specific trajectory underdetermined. High-salience events, clear anchor structures, and strong kinetic signals narrow $\mathcal{A}$ considerably, producing near-universal gaze behavior. Low salience, weak anchors, and ambiguous kinetics widen $\mathcal{A}$, producing divergent viewer trajectories. The filmmaker manages the breadth of $\mathcal{A}$; the viewer traverses it.

The connection to ecological perception is worth noting here. Gibson's affordances [9] are the action possibilities that an environment offers an organism — what it can do from a given position. In the present framework, $\mathcal{A}(x)$ is precisely the set of admissible continuations from state x: the afforded trajectories. The filmmaker engineering $(\mathcal{X}, \mathcal{A})$ is constructing an environment of affordances for the viewer's attention. Level designers have long operated with this intuition implicitly; the traversal framework makes it explicit and connects it to the formal theory of admissibility structures.

4 Kinetic Synchronization

4.1 Spatial and Temporal Guidance Distinguished

Sections 2 and 3 have addressed a spatial question: given a frame, where can attention go? The gaze field t, the salience function w, and the anchor relation A_G all concern the distribution of possible fixations across the two-dimensional plane of the screen.

But guiding attention through space is only half the filmmaker's task. There is a second, orthogonal dimension of control: the temporal structure of when attention should move, how rapidly transitions should occur, and what rhythm of change the viewer's perceptual system should be entrained to follow. This

temporal dimension is the subject of the present section, and it is here that the connection between cinema, music, and game design becomes most explicit.

The key observation is that temporal guidance is a separable system from spatial guidance. Music provides the clearest demonstration. A musical phrase has no screen position. It specifies no fixation point, creates no gaze anchor, and shapes no salience field. Yet it powerfully organizes temporal expectation: the listener anticipates the next beat, the resolution of a dissonance, the return of a theme. Music is pure temporal guidance operating without any spatial component. This makes it a useful limiting case. By examining what music and cinema share despite their obvious differences, we can isolate the temporal guidance system that cinema deploys alongside its spatial system.

Definition 4.1 (Trajectory Rate of Change)**.** Let $\Gamma(t)$ denote the admissibility structure at time t. The trajectory rate of change is:

$$\Lambda(t) = \left| \frac{d\Gamma}{dt} \right| \tag{14}$$

measuring how rapidly the set of admissible continuations is being restructured. High $\Lambda(t)$ corresponds to rapid perceptual change; low $\Lambda(t)$ corresponds to stasis or slow development.

Both a musical crescendo and a camera dolly increase $\Lambda(t)$. Both a sustained note and a long static take hold $\Lambda(t)$ near zero. The claim of this section is that these are equivalent operations in different representational media: both shape the temporal geometry of the observer's trajectory through the admissibility structure.

Proposition 4.2 (Rhythmic Equivalence)**.** *Let $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_2, \mathcal{A}_2)$ be traversal systems in different modalities with respective admissibility-rate profiles $\Lambda_1(t)$ and $\Lambda_2(t)$. If*

$$\Lambda_1(t) = \Lambda_2(t) \quad \textit{for all } t, \tag{15}$$

then the two systems generate equivalent temporal guidance structures, independently of modality. Any observer traversing either system experiences the same temporal shape of constraint change.

Proof. Temporal guidance is defined entirely in terms of $\Lambda(t)$: the rate at which the admissibility structure is being deformed. Two systems with identical $\Lambda(t)$ profiles impose identical rates of structural change on any observer trajectory, regardless of whether the underlying state space is visual, auditory, spatial, or logical. The modality determines the substrate of $\mathcal{X}$ and the sensory channel through which $\mathcal{A}$ is communicated to the observer; it does not enter the definition of $\Lambda(t)$. □

Proposition 4.2 licenses the cross-modal claims of this section. A slow camera dolly and a musical crescendo do not merely feel similar. When their $\Lambda(t)$ profiles match, they are formally equivalent temporal guidance operations. This is what allows a film score to substitute for, reinforce, or contradict the temporal guidance provided by the image track: the two systems are operating on the same quantity.

The degree of alignment between image and music can be quantified as an energy functional.

Definition 4.3 (Synchronization Energy)**.** Let $\Lambda_I(t)$ and $\Lambda_M(t)$ denote the admissibility-rate profiles of the image and music tracks respectively. The *synchronization energy* is

$$E_{\text{sync}} = \int_0^T \big(\Lambda_I(t) - \Lambda_M(t)\big)^2 \, dt. \tag{16}$$

$E_{\text{sync}} = 0$ corresponds to perfect synchronization; large E_{sync} corresponds to strong desynchronization.

Proposition 4.4 (Resonance Principle)**.** *Perceived kinetic synchronization between image and music is maximized when $E_{\text{sync}} \to 0$, and perceived tension between the two tracks is monotone increasing in E_{sync}. Synchronization and desynchronization are therefore a single variational parameter available to the filmmaker.*

Proof. By Proposition 4.2, two tracks with identical $\Lambda(t)$ profiles generate equivalent temporal guidance. When image and music coincide ($E_{\text{sync}} = 0$) the two guidance systems are aligned and reinforce each other with no perceptual tension between them. When they differ, the observer's temporal expectation system receives conflicting inputs from two channels; the conflict magnitude grows

monotonically with the pointwise divergence $|\Lambda_I(t) - \Lambda_M(t)|$, and the integrated conflict is E_{sync}. □

The Resonance Principle converts the filmmaker's choice of music-to-image synchronization into a variational problem: minimize E_{sync} for reinforcement; maximize it for tension, irony, or dissociation. The expressively neutral choice is not zero synchronization energy but the energy that best serves the moment's intended affective geometry.

4.2 The Cut as Beat

The cut is not only a spatial discontinuity event, as analyzed in Section 3.3. It is also a temporal event: a discrete pulse in the flow of the film. When cuts occur at regular intervals, they establish a meter. When cut frequency increases, tempo accelerates. When cuts cluster and thin out, they produce rhythmic phrasing.

This is not metaphorical. Editing rhythm has measurable effects on arousal and attentional engagement that are independent of the semantic content of the shots being joined. An action sequence cut at high frequency produces physiological arousal not because the images depict dangerous events but because rapid perceptual discontinuities — rapid spikes in $\Lambda(t)$ — prime the nervous system for response. The same images cut slowly produce a different physiological state. The rhythm is operating on the body before the content reaches interpretation.

The equivalence with music is structural. A sequence of cuts at regular intervals is a sequence of temporal events with a defined period, exactly like a musical pulse. Irregular cuts produce syncopation. A sudden acceleration of cut rate is a rhythmic intensification. A cut held back past its expected moment produces the same effect as a delayed beat: tension sustained by withholding the anticipated event.

4.3 Camera Movement as Phrase

Where cuts produce discrete temporal events, camera movements produce continuous temporal shapes. A camera move has onset, duration, and resolution,

exactly as a musical phrase does. The structural parallels are close enough to warrant explicit mapping.

A slow push-in toward a subject develops over time, building gradually, and resolves when the camera reaches its endpoint or cuts away. This is the temporal shape of a sustained tone moving toward a crescendo: continuous increase in $\Lambda(t)$ culminating in a release. The emotional correlate — a growing sense of attention, of approach, of something about to be revealed — is a product of this temporal shape, not of the content being approached.

A whip pan is a percussive accent: a brief, violent spike in $\Lambda(t)$ followed by immediate return to a new stable state. It marks a moment without developing it, exactly as a cymbal crash marks a rhythmic position without sustaining a melody.

A long take with a static or slowly moving camera holds $\Lambda(t)$ near zero for an extended period. This is the temporal equivalent of a sustained note or a fermata: a deliberate withholding of change that produces tension through its very stillness. The viewer's anticipation of future change accumulates during the hold, so that when motion or a cut finally arrives, its effect is amplified by the stored expectation.

4.4 Object Movement as Kinetic Event

Camera movement is not the only source of temporal structure within a shot. Objects moving within the frame produce kinetic events that are rhythmic before they are semantic. An actor crossing from one side of the frame to the other has a duration, a pace, and a moment of arrival. A vehicle accelerating through the frame has an onset and a peak. A door closing is a brief event with a definite endpoint.

These movements are temporal events in the admissibility structure before they acquire meaning. The director who blocks an actor to cross the frame at a particular moment is making a rhythmic decision: the crossing punctuates the scene, divides it into before and after, and produces a change in $\Lambda(t)$ that the viewer's perceptual system registers as a beat. The semantic content of the crossing — what it means for the character to move, where they are going, what

the movement implies about their state — is layered onto this kinetic structure after the fact.

This priority of kinetic structure over semantic content is one of the framework's stronger claims. It implies that experienced directors and choreographers are managing temporal geometry first and meaning second: they feel when a scene needs movement before they know why the character would move.

4.5 Synchronization and Desynchronization

Film combines spatial and temporal guidance simultaneously. When the temporal structure of image editing aligns with the temporal structure of musical phrasing — cuts falling on beats, camera moves matching melodic phrases, object movements synchronizing with rhythmic accents — the two guidance systems reinforce each other. The viewer experiences a unified kinetic field in which visual and auditory temporal structures are coherent.

When the two systems diverge, the viewer experiences their competition. Slow, lyrical music placed over rapid cutting creates a dissonance between the arousal level primed by image rhythm and the expectation structure primed by musical phrase. This dissonance is not merely unpleasant. It is expressive: the gap between the two temporal geometries can be used to produce irony, dissociation, or the particular horror of violence shown tenderly. The expressive effect arises from the structural mismatch, not from the content of either track alone.

Eisenstein's theory of montage, which argued that meaning emerges from the collision of images rather than their continuity, can be restated geometrically within this framework. Two admissibility structures with different temporal geometries, placed in rapid alternation, produce a combined structure whose properties neither possesses alone. The collision is a composition of incompatible $\Lambda(t)$ profiles, and the meaning that emerges is a property of that composition. The framework does not supersede Eisenstein's insight; it provides the machinery that explains why the collision works at the perceptual level before it reaches interpretation.

4.6 Game Pacing as Kinetic Design

Game design has developed an extensive practical vocabulary for managing temporal guidance, typically under the heading of pacing. The rhythm of combat encounters, the placement of moments of environmental stillness, the sequencing of high- and low-intensity zones through a level: these are kinetic decisions operating on $\Lambda(t)$ across a much longer time horizon than a single film shot, but with the same structural logic.

A well-paced game level produces a temporal profile in which high-$\Lambda(t)$ moments — combat, environmental hazards, time-pressured navigation — alternate with low-$\Lambda(t)$ intervals that allow the nervous system to return to baseline before the next intensification. The alternation is not merely aesthetic. It is physiological: sustained high $\Lambda(t)$ produces fatigue; sustained low $\Lambda(t)$ produces boredom. Effective pacing manages the temporal derivative of admissibility change to keep the observer in a productive range of engagement.

This is identical in structure to the compositional problem faced by a film editor managing the rhythm of a feature, or a composer managing the arc of a symphony. The substrate differs — spatial navigation versus screen viewing versus auditory experience — but the temporal geometry being managed is the same object in each case.

5 Affective Modulation

This section makes a claim that must be stated carefully to avoid misreading. We are not proposing a theory of emotion. We are proposing that certain recurring emotional effects in cinema are downstream consequences of the geometric operations described in Sections 2 through 4, and that identifying the geometric substrate of these effects is more precise and more useful than describing them in purely affective terms. The section is deliberately short. The framework is not competing with affective neuroscience. It is contributing to film theory.

5.1 Gaze Control as Precondition

The most basic claim of this section is also the most easily overlooked: you cannot feel what you have not been guided to see. Emotional response in cinema is not generated by the viewer independently of the film. It requires that the viewer's gaze has been directed to particular regions of the frame, held there for sufficient duration, and prepared by the salience and trajectory structures established in preceding shots.

This is not trivially true. It implies that the emotional architecture of a scene is determined before any narrative content is registered. A director who wants the viewer to feel fear upon seeing an object must first ensure that the object has been fixated with sufficient attention, at the right moment, against the right kinetic context. The emotional effect is not a response to the object. It is a response to a prepared trajectory encountering that object at a particular position within the admissibility structure. Two viewers who fixate different regions of the same frame at the moment of a revelatory cut will have different emotional responses not because they interpret the film differently but because their gaze trajectories through (t, w_t) have placed them in different states prior to the cut.

5.2 Kinetic Rhythm as Emotional Priming

The temporal geometry of Section 4 has affective consequences that precede semantic interpretation. Editing rhythm, camera movement, and object motion modify $\Lambda(t)$ — the rate of change of the admissibility structure — and through it the physiological state of the viewer. Heart rate, skin conductance, and muscle tension respond to $\Lambda(t)$ before the viewer has processed what is happening on screen. This is not a deficiency of the perceptual system. It is the system doing its job: preparing the body for action based on the temporal structure of incoming stimulation before the cost of full interpretation has been paid.

The implication for cinema is that the body is already in an emotional state by the time it knows why. A sequence of rapidly increasing cut frequency raises arousal and primes vigilance before the viewer consciously registers that something threatening is approaching. A long static take lowers arousal and produces a kind of stillness in the viewer that renders the subsequent event, what-

ever it is, more impactful. The kinetic structure is writing to the body in a language that operates below the threshold of narrative comprehension.

5.3 Recurring Emotional Operations

Several of the most familiar emotional effects in cinema can be described precisely as geometric operations on the gaze field and the admissibility structure. We note five without claiming completeness.

Dread. Progressive narrowing of the admissible gaze field combined with increasing $\Lambda(t)$. The viewer's trajectory is being constrained toward something they do not yet see. Salience concentrates. Alternatives disappear. The path narrows.

Relief. Re-expansion of the admissible gaze field after a period of constraint. The salience distribution widens. Alternative fixation targets reappear. $\Lambda(t)$ returns toward baseline. This is geometrically the reversal of dread, which is why relief and dread are phenomenologically complementary rather than merely oppositely valenced.

Surprise. Rapid redistribution of w_t to a region of the gaze field that was previously low-salience. The viewer's current fixation is rendered unimportant in a single frame, and a new attractor emerges where none existed. This is distinct from the cut, which resets the entire field. Surprise operates within a continuous shot by violating the local salience prediction.

Wonder. Sudden expansion of t beyond the viewer's prior bounds. A wide establishing shot after a sequence of close-ups, a revealed scale, an unexpected environment: these expand the fixation field rather than concentrating it. The viewer's gaze has nowhere adequate to settle because the field is suddenly larger than their current attention can span.

Suspense. Sustained constraint of the admissible trajectory with a visible but inaccessible resolution region. The viewer can see where the trajectory must eventually go. The admissibility structure does not yet permit arrival there. $\Lambda(t)$ is held elevated by the maintained constraint. Suspense is not uncertainty about

the destination. It is a topological condition: the destination is visible, the path to it is blocked, and the blockage is being deliberately sustained.

Definition 5.1 (Admissibility Jerk)**.** The *admissibility jerk* at time t is

$$J(t) = \left| \frac{d\Lambda}{dt} \right| \tag{17}$$

the rate of change of the trajectory rate of change. High $J(t)$ corresponds to a sudden intensification or sudden cessation of kinetic activity. Jump scares, sudden revelations, and the moment of violence after prolonged dread share the structure of a large positive spike in $J(t)$. The sudden cut to silence after intense action shares the structure of a large negative spike. Both are abrupt changes in the rate of admissibility deformation, and both produce strong physiological responses for the same structural reason.

The duality between dread and relief noted above can be stated as a theorem that also covers the remaining cases.

Theorem 5.2 (Constraint–Affect Correspondence)**.** *Let $V_A(t) = \mathrm{Vol}(\mathcal{A}(t))$ denote the volume of the admissibility structure at time t. Then:*

(i) *$dV_A/dt < 0$ (progressive constraint) corresponds to the dread family of affects: narrowing, approach, inevitability.*

(ii) *$dV_A/dt > 0$ (progressive release) corresponds to the relief family: opening, escape, resolution.*

(iii) *$dV_A/dt \approx 0$ with V_A small corresponds to suspense: sustained constraint without progression.*

(iv) *Large $|dV_A/dt|$ (high J) corresponds to shock, surprise, and sudden revelation, regardless of sign.*

Consequently, dread and relief are geometrically dual processes related by reversal of the sign of dV_A/dt, which is why they are phenomenologically complementary.

Proof. The affects in (i)–(iv) are defined operationally by their geometric descriptions in the preceding subsection. Statement (i) follows from the dread description: the admissible gaze field narrows, which is $dV_A/dt < 0$. Statement (ii)

follows from the relief description: the field re-expands, which is $dV_A/dt > 0$. Statement (iii) follows from the suspense description: $\Lambda(t)$ is elevated but V_A does not decrease further, placing the trajectory in a sustained constrained state. Statement (iv) follows from the jerk definition: rapid change in Λ implies rapid change in V_A, and the physiological signature of shock is substrate-independent. The duality claim in the corollary follows immediately from (i) and (ii): dread is $dV_A/dt < 0$ and relief is $dV_A/dt > 0$, which are sign reversals of the same quantity. □

The signed derivative of V_A is informative but treats large and small volumes symmetrically. Defining an affective potential removes this symmetry and gives a scalar quantity whose gradient directly tracks the phenomenological weight of constraint.

Definition 5.3 (Affective Potential)**.** The *affective potential* at time t is

$$\Phi_A(t) = -\log V_A(t). \tag{18}$$

Φ_A is large when the admissibility volume is small (high constraint) and small when the volume is large (low constraint).

Proposition 5.4 (Potential–Affect Correspondence)**.** *Dread corresponds to increasing affective potential:*

$$\frac{d\Phi_A}{dt} > 0. \tag{19}$$

Relief corresponds to decreasing affective potential:

$$\frac{d\Phi_A}{dt} < 0. \tag{20}$$

The rate of change $|d\Phi_A/dt|$ measures affective intensity; the sign measures valence.

Proof. $\Phi_A(t) = -\log V_A(t)$ implies $d\Phi_A/dt = -(1/V_A)\, dV_A/dt$. Since $V_A > 0$, the sign of $d\Phi_A/dt$ is opposite to the sign of dV_A/dt. By Theorem 5.2, $dV_A/dt < 0$ corresponds to dread and $dV_A/dt > 0$ to relief. Therefore $d\Phi_A/dt > 0$ corresponds to dread and $d\Phi_A/dt < 0$ to relief. The logarithm introduces a nonlinearity that amplifies changes when V_A is small: a unit decrease in admissibility volume

produces a larger change in Φ_A when the volume is already small than when it is large, which is consistent with the observation that additional constraint feels more intense when escape options are already limited. □

6 Empirical Predictions

1. **Fixation convergence.** Fixation convergence time after a cut is monotonically decreasing in gaze anchor density $|A_G|$. Testable with eye-tracking.

2. **Editing grammar emergence.** Classical editing rules (180-degree rule, eyeline matching, 30-degree rule) should be derivable as approximate solutions to min $C_{\text{reacquisition}}$. Film grammar becomes partially derivable from perceptual optimization rather than historical convention.

3. **Kinetic-affective correlation.** Emotional response intensity correlates with kinetic rhythm congruence between image and sound, measurable via skin conductance and heart rate, independent of semantic content.

4. **Game engagement.** Level completion rates and reported engagement correlate with salience field coherence and trajectory constraint quality, independent of narrative content.

5. **Traversal reward independence.** The reward signal associated with successful traversal is partially independent of novelty. Replay, rewatching, rereading, and rehearsal are predicted rather than anomalous.

Several of the predictions above follow as theorems from the framework rather than as standalone conjectures. The most directly falsifiable is the following.

Theorem 6.1 (Convergence Prediction)**.** *If* $S(t) \to 0$ *then* $\text{Var}[\gamma(t)] \to 0$*. That is, highly focused compositions must produce gaze convergence across viewers; as the salience entropy of a frame approaches zero, inter-viewer variance in fixation location approaches zero.*

Proof. By Theorem 2.5, in the limit $D \to 0$ the gaze density $\rho(g, t)$ concentrates on the maxima of w_t. As $S(t) \to 0$ the salience distribution becomes maximally

concentrated, with all mass on a single point g^*. In this limit $w_t(g^*) = 1$ and $w_t(g_i) = 0$ for $g_i \neq g^*$. The gaze density ρ therefore converges to a point mass at g^*, giving $\text{Var}[\gamma(t)] = \mathbb{E}[(\gamma(t) - g^*)^2] \to 0$. □

Theorem 6.1 converts the claim that focused compositions control attention into a directly eye-trackable prediction. Inter-viewer gaze variance is measurable from eye-tracking data, and salience entropy $S(t)$ is computable from any salience model applied to the frame. The theorem predicts a monotone relationship between the two quantities across frames, testable without reference to narrative content or viewer interpretation.

6.1 Film, Games, Music, and Proofs

The preceding sections have developed a framework for cinema in terms of gaze fields, anchor relations, and kinetic synchronization. Before presenting empirical predictions, it is worth stepping back to ask what kind of thing this framework actually describes. The answer turns out to be more general than cinema, and the generality is not incidental. It is what justifies the formal apparatus.

Definition 6.2 (Traversal System)**.** A *traversal system* is a triple $(\mathcal{X}, \mathcal{A}, \gamma)$ where $\mathcal{X}$ is a state space, $\mathcal{A} : \mathcal{X} \to 2^{\mathcal{X}}$ specifies admissible successor states at each point, and $\gamma : \mathbb{N} \to \mathcal{X}$ is a trajectory generated by an observer satisfying $\gamma(t+1) \in \mathcal{A}(\gamma(t))$. The designer constructs $(\mathcal{X}, \mathcal{A})$. The observer generates γ. Experience emerges from properties of γ rather than from the terminal state $\gamma(T)$ alone.

The symbol $\mathcal{A}$ carries three related but distinct uses throughout the paper, which we collect here for clarity. As a *relation*, $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{X}$ is the set of admissible transitions: $(x, y) \in \mathcal{A}$ if and only if y is an admissible successor of x. As a *function*, $\mathcal{A}(x) = \{y \in \mathcal{X} : (x, y) \in \mathcal{A}\}$ is the local possibility set at x — the set of states reachable in one step. As a *volume*, $V_A(x) = |\mathcal{A}(x)|$ (or, in continuous settings, the measure of $\mathcal{A}(x)$) quantifies the degree of local constraint: high V_A corresponds to wide-open possibility, low V_A to strong constraint. These three presentations are equivalent; the paper moves between them according to whether the argument concerns structure (the relation), local choice (the function), or intensity

of constraint (the volume). The affective potential $\Phi_A(t) = -\log V_A(\gamma(t))$ introduced in Section 5.3 is the logarithm of the local volume along the realized trajectory.

The filmmaker and the game level designer are both constructing traversal systems. The filmmaker controls a visual field; the level designer controls a navigable volume. But the mathematical object they each produce is the same: an admissibility structure from which observer trajectories emerge. The correspondences between cinematic technique and game design follow as a corollary of this shared structure rather than as an analogy between superficially similar practices.

Film	**Game**
Tracking shot	Guided movement corridor
Rack focus	Highlighted interactable
Eyeline match	Navigation marker
Lighting cue	Environmental affordance
Cut	Teleportation / loading transition
Montage sequence	Fast travel / compressed traversal

Table 1: Cinematic techniques and their game design counterparts as instances of the same operation on $(\mathcal{X}, \mathcal{A})$ in different traversal media.

Each row of Table 1 describes an identical operation applied to a different substrate. A tracking shot and a guided movement corridor are both mechanisms for imposing a constrained path through $(\mathcal{X}, \mathcal{A})$: one by controlling what enters the frame, the other by controlling where the body can go. A rack focus and a highlighted interactable both perform salience concentration: one by optical means, the other by visual marking. The difference between film and game, at this level of description, is not structural. It is the dimensionality of the state space being traversed and the degree of agency granted to the observer.

The framework extends beyond visual media. Consider a composer constructing a harmonic sequence. The state space $\mathcal{X}$ consists of tonal positions; $\mathcal{A}(x)$ specifies which harmonic moves are admissible from the current position given the established key, meter, and phrase structure. The listener generates a trajectory through expectation: at each moment, the heard sequence either satisfies or violates the locally predicted continuation. The composer does not con-

trol what the listener expects. The composer constructs the admissibility structure from which expectation trajectories emerge. Tension, resolution, surprise, and satisfaction are properties of γ relative to $\mathcal{A}$, not properties of individual notes.

The most surprising instance, and therefore the most theoretically important, is the mathematical proof.

6.2 Mathematical Proofs as Traversal Systems

A proof has no visual salience requirement, no camera, no narrative, no actor, and no soundtrack. If the traversal framework applies to proofs, it cannot be secretly about any of these things. The proof case is the stress test that separates a theory of media from a genuine theory of traversal.

When a mathematician writes a proof, they construct a traversal system in which $\mathcal{X}$ is a space of propositions and logical states, and $\mathcal{A}(x)$ specifies the inferential moves admissible from position x given the established axioms, definitions, and previously derived lemmas. The reader generates a trajectory through this space: each step either follows admissibly from the preceding position or it does not. The theorem is the terminal state. But the theorem is not the point.

This claim requires emphasis because it is counterintuitive. A reader who already knows the theorem still reads the proof. A mathematician re-reads proofs they have verified many times. If the value of a proof resided entirely in its conclusion, this behaviour would be irrational. It is not irrational. What the reader is doing is traversing the admissibility structure, and that traversal is itself rewarding. The experience of understanding a proof — of following each step as it falls into place, of feeling the argument tighten as the terminal state approaches — is an experience of trajectory properties, not of propositional content.

The proof case establishes three things that cinematic and musical examples alone cannot. First, traversal reward R_T is independent of perceptual modality: it arises in a domain that is purely logical. Second, the admissibility function $\mathcal{A}$ need not be continuous or spatial: inferential admissibility is discrete and propositional. Third, the distinction between designer and observer is not an artifact of media authorship but a structural feature of any system where one

agent specifies transition rules and another generates a path through them. A proof's author and a proof's reader stand in exactly the Designer–Observer relationship formalized in Theorem 7.1.

Gaze is the entry point of the paper because it is measurable. The proof case is the entry point to the general theory because it demonstrates that the object is traversal rather than perception. The meaning $\not\Rightarrow$ visuality argument is complete: a structure with no visual component whatsoever instantiates the same $(\mathcal{X}, \mathcal{A}, \gamma)$ triple as cinema, game design, and music.

The cross-domain correspondences in Table 1 and the extensions to music and proofs are not analogies drawn between superficially similar activities. They follow from a single structural fact: all of these systems are traversal systems in the sense of the definition above. We state this as the paper's central theorem.

Theorem 6.3 (Traversal Equivalence)**.** *Let $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_2, \mathcal{A}_2)$ be two designed admissibility structures. Suppose there exists a structure-preserving map $\phi : \mathcal{X}_1 \to \mathcal{X}_2$ such that $\phi(\mathcal{A}_1(x)) = \mathcal{A}_2(\phi(x))$ for all $x \in \mathcal{X}_1$. Then for any trajectory γ_1 through $(\mathcal{X}_1, \mathcal{A}_1)$, the image $\phi \circ \gamma_1$ is a trajectory through $(\mathcal{X}_2, \mathcal{A}_2)$, and the two trajectories share identical traversal properties: local constraint structure, admissibility density, and continuity across discontinuities.*

Proof. If $\gamma_1(t+1) \in \mathcal{A}_1(\gamma_1(t))$ then $\phi(\gamma_1(t+1)) \in \phi(\mathcal{A}_1(\gamma_1(t))) = \mathcal{A}_2(\phi(\gamma_1(t)))$, so $\phi \circ \gamma_1$ satisfies the admissibility constraint in $(\mathcal{X}_2, \mathcal{A}_2)$. Traversal properties are defined in terms of $\mathcal{A}$ alone; since ϕ is structure-preserving, these are invariant under ϕ. □

Theorem 6.3 is what converts Table 1 from an observational list into a structural result. A tracking shot and a guided movement corridor are not merely similar-looking devices. They are the same operation on isomorphic admissibility structures instantiated in different state spaces. The cross-domain transfer of design intuition that experienced filmmakers and level designers report — the sense that the problems are the same despite the substrates being different — is a recognition of this isomorphism, not a loose metaphor.

6.3 Prior Traditions Located

Several established research traditions can be located within the traversal framework as investigations of one component of the triple $(\mathcal{X}, \mathcal{A}, \gamma)$ while holding the others fixed or treating them as background.

Tradition	**Primary object**
Narrative theory	Meaning at $\gamma(T)$
Attention research	Local fixation $\gamma(t)$
Reinforcement learning	Reward $R(\gamma)$
Film grammar	Constraints on $\mathcal{A}$
Game design	Structure of $(\mathcal{X}, \mathcal{A})$
Music cognition	Temporal shape of γ

Table 2: Prior traditions as investigations of components of the traversal triple $(\mathcal{X}, \mathcal{A}, \gamma)$.

Narrative theory is primarily concerned with the meaning produced at the end of a trajectory: what does the story mean, what does the film say, what is the work about. Attention research is concerned with individual fixations: where does the eye go, what drives saccadic selection, how does salience interact with task demands. Reinforcement learning formalizes the reward signal and asks how agents learn to generate trajectories that maximize it, but typically treats $(\mathcal{X}, \mathcal{A})$ as given rather than as a design object. Film grammar — the accumulated practical knowledge of editors, cinematographers, and directors — describes constraints on $\mathcal{A}$ without formalizing either the state space or the trajectory generation process. Game design is the field that comes closest to treating $(\mathcal{X}, \mathcal{A})$ as its primary design object, but it has not generally connected its insights to the perceptual and cognitive literature on attention and gaze.

None of these traditions is wrong. Each has generated real knowledge about a genuine component of the system. What the traversal framework offers is not a replacement for any of them but a common substrate that shows where they stand relative to each other and why certain questions have fallen between them.

Recent psychoanalytic approaches locate the power of cinema in its capacity to expose structures of desire and subjectivity [24]. That project is orthogonal to the one undertaken here. Rather than beginning from desire, identification, or unconscious structure, the traversal framework begins from admissibility, tra-

jectory generation, and guided traversal. Psychoanalytic interpretation may describe one class of effects generated by cinematic trajectories; the trajectory itself is the more primitive object and is available to analysis without commitment to any particular theory of the subject.

6.4 Traversal as Intrinsically Rewarding

The traversal framework implies a non-obvious claim about the structure of reward in designed experience. Let $R(\gamma)$ denote the total reward associated with a traversal, and let $R(\gamma(T))$ denote the reward associated with the terminal state alone. The framework predicts:

$$R(\gamma) \neq R(\gamma(T)) \tag{21}$$

That is, the value of a traversal is not reducible to the value of its outcome. The path has properties that matter independently of where it ends.

This prediction is falsified by a large class of everyday observation if equation (21) is false. People replay games whose endings they know. They rewatch films whose plots hold no surprises. They reread novels, re-listen to familiar music, rehearse performances they have already given, solve puzzles whose solutions they remember, and replay famous chess games whose outcomes are recorded. Under a purely terminal-state account of reward, each of these behaviours is anomalous: the agent already possesses the information content of the endpoint, so repetition should yield no marginal value.

Under the traversal account, none of these behaviours is anomalous. The agent is not seeking information about the terminal state. The agent is re-traversing the admissibility structure $(\mathcal{X}, \mathcal{A})$, and that traversal is itself the source of reward. The reward signal is generated by properties of γ — its pacing, its local uncertainties, its moments of constraint and release, its kinetic shape — rather than by the information content of $\gamma(T)$.

This is consistent with the literature on intrinsic motivation and flow. Csikszentmihalyi's account of optimal experience [33] identifies a state of absorbed engagement arising when challenge and skill are matched — which in traversal terms corresponds to a trajectory that is neither trivially unconstrained nor

impossibly narrow. Koster's theory of game enjoyment [34] argues that fun emerges from traversing and mastering structured possibility spaces, which is precisely the traversal account stated informally. The intrinsic reward literature [40, 41, 42] establishes that agents derive reward from the properties of information flow during exploration, independent of terminal outcomes. The traversal framework gives these observations a common formal substrate.

This observation reaches beyond aesthetics. Ritual, rehearsal, education, and many forms of play may all derive part of their value from the intrinsic properties of structured traversal rather than from the informational or instrumental value of their endpoints. The geometry of the path is not merely a means to the destination. In many of the most important human activities, the geometry of the path is the point.

The inequality $R(\gamma) \neq R(\gamma(T))$ can be strengthened into a decomposition theorem.

Theorem 6.4 (Traversal Reward Decomposition)**.** *There exists a functional R_T on trajectories such that*

$$R(\gamma) = R\big(\gamma(T)\big) + R_T(\gamma), \tag{22}$$

where $R_T(\gamma)$ is the traversal reward — the component of total reward attributable to properties of the path rather than the endpoint. The traversal reward $R_T(\gamma)$ vanishes if and only if $(\mathcal{X}, \mathcal{A})$ is a degenerate admissibility structure in which every path to the terminal state has identical local properties.

Proof. Define $R_T(\gamma) = R(\gamma) - R(\gamma(T))$. This is a functional on the trajectory that captures all reward not attributable to the terminal state; equation (22) holds by construction. It remains to show that $R_T \not\equiv 0$ for non-degenerate $(\mathcal{X}, \mathcal{A})$. The empirical evidence is extensive: replay of games, films, music, proofs, and ritual demonstrates that $R(\gamma) > R(\gamma(T))$ for known trajectories, establishing $R_T > 0$. Conversely, if all paths to $\gamma(T)$ in $(\mathcal{X}, \mathcal{A})$ share identical local properties — that is, if $\mathcal{A}$ allows only one effective trajectory — then path properties cannot vary and $R_T = 0$. Non-trivial $\mathcal{A}$ therefore generically produces non-zero R_T. □

The decomposition (22) makes the paper's central claim precise. Designed experience is not the delivery of a terminal state to the observer. It is the con-

struction of an admissibility structure $(\mathcal{X}, \mathcal{A})$ whose traversal generates $R_T > 0$ independently of what $\gamma(T)$ contains. The designer who understands this optimizes the path, not merely the destination.

6.5 Case Study: *Perfect Days* and Reward Without Resolution

Most films conceal the traversal reward within a terminal reward structure: the plot moves toward a resolution, the trajectory matters because it leads there, and R_T is entangled with $R(\gamma(T))$ in ways that are difficult to disentangle analytically. Wim Wenders' *Perfect Days* [13] (2023) removes this entanglement almost entirely, providing what amounts to a natural experiment for the Traversal Reward Decomposition.

The film follows Hirayama, a Tokyo toilet cleaner, through a routine that varies only within narrow bounds. He wakes, washes, drives, cleans, eats, reads, photographs trees, and sleeps. The next day he does the same. No mystery is solved. No antagonist is defeated. No transformation occurs. Minimal narrative progression means that conventional terminal-state reward is small relative to what most films accumulate across their running time. A more defensible formulation than $R(\gamma(T)) \approx 0$ is the dominance claim:

$$R_T(\gamma) \gg R\big(\gamma(T)\big). \tag{23}$$

The final scene does contain genuine terminal effects — emotional integration, retrospective recontextualization, character recognition — so the terminal term is not zero. What the film demonstrates is that these terminal effects are dwarfed by the accumulated traversal reward, and that the terminal effects themselves depend causally on the prior traversal: without γ, $R(\gamma(T))$ collapses too. The film is therefore a case where traversal reward dominates, and where the terminal reward is itself a function of the path rather than independent of it.

Repetition as traversal stabilization. Most narrative theories predict that repeated events should become less engaging because they reduce information gain. Yet the film presents the same sequence — waking, washing, driving, cleaning, lunch, reading — across multiple days with only minor variation, and does not lose the viewer. The traversal framework predicts this. The admis-

sibility structure $(\mathcal{X}, \mathcal{A})$ is approximately stable across days, but the realized trajectory γ_i differs from γ_j at the level of local variation: the quality of light on a particular morning, a brief encounter, a slightly different lunch. The viewer is not tracking outcomes. The viewer is tracking variation within constraint, which is precisely the regime in which $R_T > 0$ independently of information at the terminal state.

The tree photographs. Hirayama repeatedly photographs sunlight through leaves. From a semantic perspective these scenes are nearly empty: no plot advances, no information is revealed. Yet they are among the film's most memorable moments. The traversal account explains this. The photographs create localized gaze attractors: the viewer's eye is drawn through a recurring micro-trajectory

$$\text{road} \;\rightarrow\; \text{tree} \;\rightarrow\; \text{light} \;\rightarrow\; \text{shadow} \tag{24}$$

whose repetition across the film establishes it as a stable attentional motif. The salience gradient ∇w_t consistently draws the eye from ambient environment to filtered light to the photographic act itself. The reward is not informational. It is the pleasure of a familiar trajectory traversed again with local variation.

The toilet architecture. The restrooms Hirayama cleans are not merely settings. They are admissibility structures. Each space constrains movement, visibility, cleaning order, and action sequence: $(\mathcal{X}_{\text{toilet}}, \mathcal{A}_{\text{toilet}})$ is a genuine traversal system with its own geometry. Hirayama traverses these structures with complete competence, and the camera presents his traversal without truncation or abbreviation. The satisfaction of watching this is closely related to the pleasure of woodworking videos, cooking videos, repair videos, and speedruns: in each case the endpoint is known or trivial, and the reward derives from observing competent traversal of a constrained space. The constraint is necessary: incompetent traversal of an unconstrained space produces no reward. R_T is generated by the interaction of a capable trajectory generator with a non-trivial $\mathcal{A}$.

Music as temporal admissibility. The cassette tapes that Hirayama plays while driving are temporally precise. Each song establishes immediately a set of expectations regarding rhythm, phrase completion, and harmonic continuation: a

temporal admissibility structure $(\mathcal{T}, \mathcal{A}_{\mathcal{T}})$ running alongside the visual one. The viewer traverses both simultaneously, and the reward of each scene in the van comes from the synchronized traversal of spatial gaze trajectory and musical expectation trajectory. This directly enacts the kinetic synchronization analysis of Section 4: when $E_{\text{sync}} \to 0$, the two $\Lambda(t)$ profiles align, and the combined traversal is more rewarding than either alone.

The final scene. The film ends on Hirayama's face during the morning drive. Nothing is resolved. No information is revealed that was not already available. Yet the scene is widely described as emotionally overwhelming. The traversal framework provides a precise account. By the time the final scene arrives, the viewer has accumulated a detailed internal model of Hirayama's admissibility structure: the bounds of his world, the consistency of his constraints, the texture of his traversals. The face is rewarding not because of information revealed at $\gamma(T)$ but because it recontextualizes the entire prior trajectory. The emotional response is a function of $R_T(\gamma)$ accumulated across the whole film, activated by the final frame. A viewer who watched only the final scene without the preceding trajectory would find it nearly empty. This is close to a direct empirical demonstration of the decomposition: remove the traversal, and the terminal state loses most of its reward. The reward was in the path.

7 Conclusion

The geometry of attention was the entry point because it is observable. The geometry of traversal is the more general object.

This paper began with a specific question about cinema: what are cinematic techniques actually doing at the perceptual level, prior to narrative interpretation? The answer developed across Sections 2 through 4 is that they are constructing and manipulating an admissibility structure $(\mathcal{X}, \mathcal{A})$ over a visual field, from which the viewer generates a gaze trajectory γ by local admissible steps. Editing grammar is not convention. It is approximate solutions to a re-acquisition minimization. Camera movement is not style. It is trajectory imposition or trajectory constraint. Musical phrasing is not accompaniment. It is temporal guid-

ance operating on the same trajectory structure as the image, through a different sensory channel.

But the framework did not stay in cinema. Section 6 showed that the same formal structure — designer constructs $(\mathcal{X}, \mathcal{A})$, observer generates γ — describes game level design, musical composition, and the writing of mathematical proofs. The proof case is the strongest argument for the framework's generality. A mathematician and a filmmaker are not doing similar things metaphorically. They are doing the same thing formally: specifying a space of states and a function over that space that determines what comes next. The reader and the viewer are not having analogous experiences. They are both traversing an admissibility structure, and in both cases the traversal is itself rewarding, independently of what the terminal state contains.

Film theory has tended to ask: what does a film mean? The traversal framework suggests a prior question: what does a film do? Before it means anything, a film constructs a sequence of fixation fields, manages anchor density across discontinuities, imposes kinetic rhythms on the viewer's body, and channels attention through a carefully controlled landscape of possibility. Meaning, where it arises, arises from trajectories through that landscape.

7.1 Traversal and Biological Cognition

The traversal framework reaches beyond aesthetic design into the architecture of biological cognition itself. Organisms do not traverse possibility spaces merely to secure rewards at terminal states. If this were so, behavior would collapse toward the shortest path to known outcomes, and exploration would cease once those outcomes became predictable. But biological systems do not behave this way.

A clarification is required before proceeding. Throughout the paper, the designer of $(\mathcal{X}, \mathcal{A})$ has been a human agent: a filmmaker, a level designer, a composer, a mathematician. In biological settings no such agent need exist. The admissibility structure that an organism navigates is constructed by evolutionary history, by developmental processes, and by the physical structure of the environment itself. A foraging animal traversing a landscape, a child explor-

ing a room, an immune system navigating antigen space: in each case $(\mathcal{X}, \mathcal{A})$ is provided not by deliberate design but by the structure of the world the organism inhabits. The traversal framework does not require intentional design. It requires only that a state space exist and that local transitions be constrained. Evolution and environment are adequate constructors of $(\mathcal{X}, \mathcal{A})$; intentional designers are a special case, and a particularly tractable one for formal analysis because their design decisions are directly observable.

Animals play after nutritional needs are satisfied. Humans revisit familiar places, rehearse known skills, replay completed games, reread books whose endings they remember, and repeatedly engage with environments containing no novel terminal information. Under a purely outcome-centered theory of reward each of these behaviours is anomalous. Under the traversal framework — and specifically under the Traversal Reward Decomposition of Theorem 6.4 — they are expected. $R_T(\gamma)$ is non-zero for non-degenerate admissibility structures regardless of how many times the structure has been traversed before.

This suggests a reinterpretation of intrinsic motivation as a preference over admissibility structures. Curiosity becomes the search for regions of state space whose local constraints are neither trivial nor chaotic. Play becomes repeated traversal of a richly structured admissibility manifold. Learning becomes the progressive expansion of the set of admissible trajectories available to the organism. These accounts align naturally with ecological theories of perception [9, 11], information foraging [43, 44], and active inference [46, 49]: all describe organisms as agents that generate trajectories through constrained environments while maintaining sufficient predictive coherence to avoid collapse.

An environment that is completely predictable provides no opportunity for adaptive traversal. An environment that is completely unpredictable prevents stable traversal altogether. Productive cognition occupies an intermediate regime in which admissibility constraints exist but are not fully exhausted — a formalization of the flow channel identified by Csikszentmihalyi [33].

From this perspective, designed experiences — films, games, musical compositions, mathematical proofs, rituals, sports, and educational systems — are artificial habitats for constrained traversal. They construct admissibility structures whose function is not merely to deliver information or narrative but to pro-

vide trajectories worth traversing. The designer is creating a temporary ecology of attention within which prediction, exploration, correction, and continuation can occur. The reward associated with these systems emerges from successful participation in a structured process of traversal that mirrors the foundational requirements of biological cognition itself.

The ontology implied by the full framework is therefore:

$$\text{Constraint} \longrightarrow \text{Trajectory} \longrightarrow \text{Experience} \longrightarrow \text{Meaning} \tag{25}$$

rather than the reverse order assumed by most interpretive traditions. Meaning appears last, not first. The admissibility structure is the primary object.

The observer experiences freedom of traversal. The designer experiences control of admissibility. The filmmaker and the level designer construct the geometry. The observer supplies the path.

Theorem 7.1 (Designer–Observer Separation)**.** *The designer specifies $(\mathcal{X}, \mathcal{A})$ but does not specify γ. The observer specifies γ but cannot escape $\mathcal{A}$: every observer-generated trajectory satisfies $\gamma(t+1) \in \mathcal{A}(\gamma(t))$ by definition of the traversal system. Consequently, designed experience is the interaction of constraint and traversal rather than the unilateral control of either.*

Proof. By the definition of a traversal system, the designer constructs $(\mathcal{X}, \mathcal{A})$ and the observer generates γ subject to $\gamma(t+1) \in \mathcal{A}(\gamma(t))$. The designer cannot specify γ because the observer's generative process is not part of the design; two observers in the same $(\mathcal{X}, \mathcal{A})$ generate different γ in general. The observer cannot escape $\mathcal{A}$ because admissible continuation requires membership in $\mathcal{A}(\gamma(t))$ at each step. The interaction is therefore irreducibly bilateral: neither party fully determines the experience alone. □

Theorem 7.1 is the formal ground for the paper's central observation. The filmmaker who believes they are controlling the viewer's experience is correct that they are controlling $\mathcal{A}$, and incorrect if they believe this determines γ. The viewer who believes they are freely choosing where to look is correct that they are generating γ, and incorrect if they believe this is independent of $\mathcal{A}$. The experience lives in the interaction between the two.

Principle of Designed Experience

Design does not determine trajectories.
Design determines admissibility.
Trajectories emerge from admissibility.
Experience emerges from trajectories.

Traversal is not merely one source of reward among others. It may be one of the fundamental organizational principles through which biological and artificial systems alike transform possibility into experience. The architectures we build do not simply communicate meanings or deliver outcomes. They shape the admissibility structures through which trajectories become possible, and through those trajectories, the forms of experience available to consciousness itself.

A Collected Formal Definitions and Results

This appendix collects all formal definitions, propositions, theorems, and corollaries from the main text.

Definitions

D1. Gaze Field (t, w_t) §2.1

D2. Salience Entropy $S(t)$ §2.1

D3. Gaze Density Field $\rho(g, t)$ §2.1

D4. Gaze Trajectory $\gamma : [0, T] \to$ §3.1

D5. Camera Motion Field $V_t : t \to Tt$ §3.2

D6. Re-acquisition Cost $C(F_t, F_{t+1})$ §3.3

D7. Gaze Anchor Relation $A_G(F_t, F_{t+1})$ §3.3

D8. Anchor Entropy $H_A(F_t, F_{t+1})$ §3.3

D9. Trajectory Rate of Change $\Lambda(t)$ §4.1

D10. Synchronization Energy E_{sync} §4.1

D11. Traversal System $(\mathcal{X}, \mathcal{A}, \gamma)$ §6.1

D12. Admissibility Jerk $J(t) = |d\Lambda/dt|$ §5.3

D13. Affective Potential $\Phi_A(t) = -\log V_A(t)$ §5.3

Propositions

P1. Salience Concentration §2.1, Prop. 2.3

P2. Local Trajectory Constraint §3.1, Prop. 3.2

P3. Piecewise Continuity §3.1, Prop. 3.3

P4. Least-Effort Saccade Principle §3.1, Prop. 3.4

P5. Trajectory Alignment §3.2, Prop. 3.6

P6. Grammar as Optimization §3.3, Prop. 3.13

P7. Rhythmic Equivalence §4.1, Prop. 4.2

P8. Resonance Principle §4.1, Prop. 4.4

P9. Potential–Affect Correspondence §5.3, Prop. 5.4

Theorems

T1. Salience Flow §2.1, Thm. 2.5

T2. Tracking Stability §3.2, Thm. 3.7

T3. Anchor Bound §3.3, Thm. 3.10

T4. Information-Theoretic Cut Cost §3.3, Thm. 3.12

T5. Constraint–Affect Correspondence §5.3, Thm. 5.2

T6. Convergence Prediction §6, Thm. 6.1

T7. Traversal Equivalence §6.1, Thm. 6.3

T8. Traversal Reward Decomposition §6.4, Thm. 6.4

T9. Designer–Observer Separation §7, Thm. 7.1

Corollaries

C1. Grammar Emergence §3.3, Cor. 3.14

C2. Traversal Reward Decomposition $R(\gamma) = R(\gamma(T)) + R_T(\gamma)$ §6.4, Thm. 6.4

The organizing diagram

$$(\mathcal{X},\, \mathcal{A}) \xrightarrow{\text{observer generates}} \gamma \xrightarrow{\text{trajectory properties}} R(\gamma) = R\big(\gamma(T)\big) + R_T(\gamma) \qquad (26)$$

The designer controls the left term. The observer controls the middle term. The right term decomposes into endpoint reward and traversal reward; for non-degenerate admissibility structures, both terms are non-zero.

References

[1] Yarbus, A. L. (1967). *Eye Movements and Vision* (B. Haigh, Trans.). Plenum Press.

[2] Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

[3] Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

[4] Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222.

[5] Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 1–23.

[6] Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28.

[7] Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections: The Journal for Movies and Mind*, *2*(1), 1–26.

[8] Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*(3), 432–439.

[9] Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.

[10] Norman, D. A. (1988). *The Design of Everyday Things*. Basic Books.

[11] Noë, A. (2004). *Action in Perception*. MIT Press.

[12] Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

[13] Wenders, W. (Dir.). (2023). *Perfect Days* [Film]. Master Mind / Spoon.

[14] Eisenstein, S. (1949). *Film Form: Essays in Film Theory* (J. Leyda, Ed. & Trans.). Harcourt Brace.

[15] Bordwell, D., Staiger, J., & Thompson, K. (1985). *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. Columbia University Press.

[16] Bordwell, D. (2008). *Poetics of Cinema*. Routledge.

[17] Thompson, K. (1999). *Storytelling in the New Hollywood: Understanding Classical Narrative Technique*. Harvard University Press.

[18] Carroll, N. (1996). *Theorizing the Moving Image*. Cambridge University Press.

[19] Plantinga, C. (2009). *Moving Viewers: American Film and the Spectator's Experience*. University of California Press.

[20] Shimamura, A. P. (Ed.). (2013). *Psychocinematics: Exploring Cognition at the Movies*. Oxford University Press.

[21] Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*(1), 1–27.

[22] Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring Cognition at the Movies* (pp. 165–191). Oxford University Press.

[23] Smith, T. J. (2017). The attentional theory of cinematic continuity revisited. *Projections*, *11*(1), 70–89.

[24] Rollins, H. (2024). *Psychocinema*. Polity Press.

[25] Ryan, M.-L. (2001). *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media*. Johns Hopkins University Press.

[26] Ryan, M.-L. (2015). *Narrative as Virtual Reality 2: Revisiting Immersion and Interactivity in Literature and Electronic Media*. Johns Hopkins University Press.

[27] Lynch, K. (1960). *The Image of the City*. MIT Press.

[28] Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press.

[29] Alexander, C. (1979). *The Timeless Way of Building*. Oxford University Press.

[30] Jenkins, H. (2004). Game design as narrative architecture. In N. Wardrip-Fruin & P. Harrigan (Eds.), *First Person: New Media as Story, Performance, and Game* (pp. 118–130). MIT Press.

[31] Nitsche, M. (2008). *Video Game Spaces: Image, Play, and Structure in 3D Worlds*. MIT Press.

[32] Totten, C. W. (2014). *An Architectural Approach to Level Design*. CRC Press.

[33] Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper and Row.

[34] Koster, R. (2013). *A Theory of Fun for Game Design*. O'Reilly.

[35] Schell, J. (2019). *The Art of Game Design: A Book of Lenses* (3rd ed.). CRC Press.

[36] Meyer, L. B. (1956). *Emotion and Meaning in Music*. University of Chicago Press.

[37] Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.

[38] London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter* (2nd ed.). Oxford University Press.

[39] Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press.

[40] Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*. McGraw-Hill.

[41] Schmidhuber, J. (1991). Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 1458–1463).

[42] Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*, 6.

[43] Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675.

[44] Pirolli, P. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press.

[45] Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

[46] Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

[47] Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

[48] Seth, A. (2021). *Being You: A New Science of Consciousness*. Dutton.

[49] Parr, T., Pezzulo, G., & Friston, K. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

[50] Milnor, J. (1963). *Morse Theory*. Princeton University Press.

[51] Spivak, M. (1979). *A Comprehensive Introduction to Differential Geometry* (2nd ed.). Publish or Perish.

[52] Gallier, J., & Quaintance, J. (2020). *Differential Geometry and Lie Groups: A Computational Perspective*. Springer.

[53] Gromov, M. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser.

[54] Bridson, M. R., & Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature*. Springer.