
Das Kinodispositiv

Claude Bailblé

Gibt es Gründe, sich erneut mit der Trägerstruktur des Kinos, dem sogenannten «Dispositiv» zu beschäftigen, ist das Thema nicht umfassend behandelt worden, ist die Diskussion nicht abgeschlossen? Doch jüngere Entwicklungen im Bereich der wissenschaftlichen Wahrnehmungstheorie drängen nachgerade darauf, das Thema noch einmal in Angriff zu nehmen, da sie die visuellen und auditiven Unterschiede zwischen Film und Wirklichkeit auf neue Weise nachvollziehbar machen: jene Unterschiede, welche das Kino zu einer Art «gesteigerten Realität» machen, die interessanter ist als der sie umgebende Alltag. Zwar mögen die Bilder und Töne eines Films denen der normalen Wahrnehmung so weit ähneln, dass sie glaubhaft wirken; doch sie unterscheiden sich auch genügend, um ein Repräsentationssystem zu konstituieren, das intensiver oder dichter ist als die gewöhnliche Realität. Das Dispositiv – das Zusammenspiel von optischen und akustischen Apparaten – erlaubt es dem Zuschauer, sich ins Schicksal der Figuren und die virtuelle Szenerie des Films hineinzuprojizieren und dabei Kinoraum und Artefaktcharakter der *Mise-en-scène* zu vergessen.

Zweifellos basiert die unmittelbare Glaubwürdigkeit der Filme auf der Ähnlichkeit der im Kinosaal frontal angeordneten Bilder und Töne mit unserer Wahrnehmung der Wirklichkeit. Man erinnere sich an Alberti (2002) und die von ihm theoretisierte «Glasscheibenperspektive» der Maler, die wie ein zur Welt geöffnetes Fenster funktioniert: Gemälde böten, wenn sie auf der gesamten Fläche in derselben Schärfe gehalten seien, wie wir sie real nur in der je zentralen Blickzone erleben,¹ den Augen die Möglichkeit, die dargestellte Szene frei zu

1 Das Objektfeld – so gut ausgeleuchtet es auch sein mag – könnte ohne Hilfe des Auges kein Bild erzeugen. Dieses Bild ist vollwertig aber nur im Bereich von 5 Grad (dem «Blickfleck»). Die Ausweitung dieser Schärfezone ist deswegen hoch artifiziell

erkunden, ganz als schaute man durch ein Fenster nach draußen. Indem er sich den Blickpunkt des Malers zu Eigen mache, zentriere der Betrachter sich spontan vor dem Gezeigten: So entstehe aus der Abtastung des Gemäldes mit den Augen eine gewisse Wirklichkeitsillusion.

Mit dem Kameraobjektiv hat sich dieses Prinzip des Quattrocento in die Fotografie hinein verlängert. Sobald das fotografische Bild komplett scharf ist, auf der gesamten Fläche sowie in der Tiefe des Sichtfelds, dominiert die im Bild gezeigte *Ansicht* und verdrängt den *Blickpunkt*, der sie – durch das Okular des Apparates – erst erzeugt hat. So treten die Wahl des Ausschnitts und ebenso die Wahl des Augenblicks der Aufnahme hinter der repräsentierten, optisch gesättigten Realität zurück, sodass der Inhalt der Bilder aufgewertet wird. Infolge dieser Umschichtung hat sich die Vorstellung etabliert, die Realität existiere latent schon immer und überall in dieser hyperscharfen Form, und es gehe nur noch darum, sie zu kadrieren, um eine bedeutsame *Ansicht*, einen vermeintlich authentischen Moment festzuhalten. Das fotografische Bild wird zu einer Art *optischen Falle* [*piège à regard*], denn es produziert den Glauben an ein Fenster, das sich auf eine vollständig lesbare Welt hin öffnet.

Im abgedunkelten Projektionsraum hat das frühe Kino diese Vorstellung noch bestärkt: Das Leinwandformat sorgte für einen räumlichen, die Dauer der Plansequenz für einen zeitlichen Ausschnitt, der seinerseits die Annahme genereller Schärfe postuliert. Die Möglichkeit, alles scharf zu filmen, arbeitet der *Fovea centralis* zu, diesem taktilen, hoch auflösenden Zentrum des begehrenden Blicks.* So wurde eine Art gesteigerten Sehens geschaffen, das geeignet ist, die Schaulust zu befriedigen. Ermöglicht wird diese Überhöhung des Sichtbaren dadurch, dass ihr instrumenteller Ursprung – Kamera und Projektor – kaschiert ist.

Von diesen Grundgegebenheiten aus entwickelten sich die verschiedenen (strukturellen und szenografischen) Arrangements, die das Kinodispositiv charakterisieren und die sich um die Aufmerksamkeit und Aufnahmefähigkeit des Zuschauers zentrieren, der gleichermaßen bewegt wie unbewegt vor der Leinwand sitzt.

(*Perspectiva artificialis*), trägt aber bei zum Glauben an eine in aller Schärfe bereits vor ihrer Aneignung durch die Augen vorhandene Welt.

* [Anm.d.Ü.:] Bei der *Fovea centralis* (auf Deutsch auch *Sehgrube* genannt), handelt es sich um den schärfsten (höchstauflösenden) Bereich auf der Netzhaut. Automatisch werden Objekte nach Möglichkeit immer so fixiert, dass ihre Abbilder genau in der *Fovea centralis* liegen.

Die Erfindung der Montage

Zwar bediente sich das Kino bei der Malerei (hinsichtlich der Perspektive und der frontalen Leinwand), bei der Fotografie (bezüglich des Schwarzweiß und der Beleuchtung), bei der Architektur (bezüglich des Dekors und des *Trompe-l'œil*) beim Theater (hinsichtlich des Spiels der Darsteller, auch wenn es beim Film nicht kontinuierlich erfolgt), beim Roman (in seinem Erzählfluss) oder sogar bei der Musik (zur emotiven Untermauerung). Doch hat es zumindest *einen* Bereich erfunden, der ihm allein gehört: die Montage. Zwischen 1905 und 1913 emanzipieren sich die Filme vom autarken Tableau der Brüder Lumière (vgl. Burch 1990), um jenen paradoxalen Ort zu erschaffen, der als Leinwand zwar stets immobil und zentriert bleibt, sich als Filmbild aber permanent verändert und damit einen provisorischen Charakter besitzt: Szenen unterschiedlicher Dauer, aus verschiedenen Distanzen und Blickachsen präsentieren sich in dem unverrückbaren, frontalen Fenster. Im Zuge dieser Entwicklung wächst die Anzahl der Schnitte, die Filmrollen werden länger, die erzählten Geschichten komplexer: Schon bald hat man die kurzen Sketche der Frühzeit hinter sich gelassen.

Nach und nach entsteht eine *Kunst der Anschlüsse*. Anschluss-Stellen, Übergänge/Sprünge von einer Einstellung zur nächsten, ermöglichen die *Ellipse* und ihre Verschleierung hinter einer «zeitlichen Quasi-Kontinuität» (bei Bewegungs-, Geschwindigkeits- und Handlungsanschlüssen) zugunsten einer verdichteten Zeit ohne offensichtliche Nahtstellen. Sie erlauben die Anordnung von aufeinander folgenden Einstellungen in eine «räumliche Quasi-Kontinuität» (beim Betreten und Verlassen des Bildes, bei Anschlüssen der Bewegungsrichtung und der Position), als ob man die natürlichen Blicke – «schau hierhin», «schau dorthin» – ersetzen könnte durch die ausdrucksstärkeren des Blickpunktes: «schau von hier aus», «schau von dort aus». Dieser Trick basiert darauf, dass man die räumliche Orientierung des Zuschauers durch ein doppeltes Arrangement täuscht:

(1) Zunächst passiv: Indem die rechteckige Leinwand den peripheren Bereich vergessen lässt, den räumlichen Kontext ins Dunkel verbannt, wird die körperliche Einbindung des Zuschauers in den repräsentierten Schauplatz de facto vermindert;² das zweidimensionale Bild lässt den

2 Da der periphere retinale Bereich (also das Körperschema) nicht mehr angesprochen ist, werden die räumlichen Orientierungen elastisch und formbar. Das ist allerdings

dreidimensionalen Raum außerhalb des eigenen Körpers verschwimmen. Allerdings entspricht das rechteckige Bild unserem Gesichtsfeld eigentlich nicht; Großaufnahmen wirken hier anders als Halbtotale; die Einbindung des Zuschauers ändert sich je nachdem, ob ein Schauspieler frontal, im Profil, als Dreiviertel-Rückenfigur oder nur teilweise zu sehen ist; flüchtige oder langsame Bewegungen beeinflussen die Verständlichkeit; die Perspektivierung schafft eine Staffelung der Objekte vom Vorder- zum Hintergrund; Blickrichtungen und Bewegungen sind beim koplanaren, zweidimensionalen Sehen weniger genau zu erkennen als in der dreidimensionalen Wirklichkeit.

(2) Aber auch aktiv: Indem es eine Situation in Bewegung zeigt, lädt das filmische Bild dazu ein, es kontinuierlich oder diskontinuierlich durch sakkadische Augenbewegungen zu verfolgen, die den Schnitten ähneln (Sprünge in den Blickpunkten). Die bei der *Mise-en-scène** getroffenen Vorkehrungen (wie 30°- und 180°-Grad-Regel, *eyeline matching* etc.) überdecken die störenden Diskontinuitäten und lassen die Brüche oder die aus ihnen resultierenden Probleme verschwinden, die der normalen Wahrnehmung fremd sind. Genau diese Regeln, die am Drehort falsch und erst auf der Leinwand richtig erscheinen, sichern die Kontinuitätsillusion. Sie sind einem langen Prozess von *trial-and-error* zu verdanken, bis sich eine gewisse narrative Transparenz herausgebildet und stabilisiert hatte.

Aus den genannten Dispositionen entstehen drei narrative Achsen: die *situative* Achse (man sieht die ganze Situation, die Interaktionen und Bewegungen der Figuren und ihre Gesten), die *personale* Achse (man nähert sich dem einen oder anderen Protagonisten und folgt ihm mit einem Schwenk oder einer Fahrt) und die *subjektive* Achse (entweder

nicht der Fall in der *Géode* (einem Imax-Kino im Pariser Parc de la Villette), wo die Sequenz-Einstellung unabdingbar ist.

* [Anm.d.Ü.:] Es ist darauf hinzuweisen, dass Bailblé den Begriff *Mise-en-scène* in seiner weitesten Bedeutung gebraucht, die alle bei der Filmproduktion notwendigen Entscheidungen umfasst (er ließe sich auch als ‚Inszenierung‘ übersetzen). In Frankreich ist diese breite Bedeutung nicht unüblich: «In den frühesten Tagen, in den Anfängen des Kinos [und heute auch normalerweise; GK] bedeutete *mise en scène* die Anordnung des visuellen Materials vor der Kamera; nur die Eingeweihten wussten, dass dieses Wort in Wirklichkeit für alle Entscheidungen des Regisseurs stand: Kameraposition, Blickwinkel, Spiel der Darsteller. Lediglich diese Eingeweihten wussten daher, dass *mise en scène* beides beinhaltet – die Geschichte, die erzählt wird, und die Art und Weise, wie sie erzählt wird» (François Truffaut, zit.n. Bordwell, David: *Visual Style in Cinema. Vier Kapitel Filmgeschichte*. Frankfurt a.M./München: Verlag der Autoren 2001, S. 153).

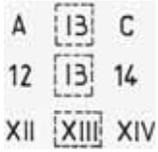
extern: die Großaufnahme eines Gesichts, aus dem man die Blicke, die Intentionen oder Reaktionen liest, oder *intern*: die subjektive Kamera, die uns mit einer Figur sehen und empfinden lässt). Seit Griffith haben die Filmemacher nie mehr aufgehört, ihre *Mise-en-scène* um diese Achsen herum zu organisieren, sei es, um die Aufmerksamkeit zu lenken (beispielsweise in der Parallelmontage: die Zuschauer wissen mehr oder weniger als bestimmte Figuren und sehen mehrere Situationen sich gleichzeitig entwickeln), sei es, um Identifikationen in Gang zu setzen (die Zuschauer erleben die Konflikte, Wünsche und Befürchtungen der verschiedenen Figuren mit).

Die Kunst des Anschlusses wird so zu einer Kunst des Ausschlusses. Denn es geht darum, die den Situationen und Akteuren jeweils zugestandene *Zeit zu hierarchisieren*, die *signifikanten Momente auszuwählen*, Hinweise ausführlich oder knapp ins Bild zu bringen oder die für die Protagonisten *entscheidenden Momente zu rhythmisieren* (um Anteilnahme oder Ablehnung auszulösen). Des Weiteren erfordert es die Dynamik der Erzählung, die situativen, personalen oder subjektiven Achsen zu alternieren, um der Qualität des Schauspiels – seiner Stimmigkeit und Emotionalität – ebenso Rechnung zu tragen wie der Dramaturgie (was darf man wissen, was soll man ignorieren, was vermuten?) und dem Generieren eines Erwartungshorizontes (durch innere Spannung der Erzählung, durch rote Fäden).

Mit dem Tonfilm ergänzt oder komplettiert das auditive Gedächtnis das visuelle. Wenn die (Halb-)Totale die Orte und Bewegungen erfasst und die Nahaufnahme die Gesichter mit den Blick- und Sprechrichtungen, so öffnen die Stimmen und Geräusche aus dem Off einen größeren szenischen Raum (und erhalten ihn aufrecht), der die gesamte Sequenz umfasst, einschließlich der Gegenschüsse, die noch kommen sollen oder die schon erfolgt sind.³ Imaginiertes, im Gedächtnis Gespeichertes und Wahrgenommenes reichen über die unmittelbare Einstellung hinaus und wirken zusammen, um die Gegenwart der Figuren aufzuladen. Die Filmemacher entdecken und gestalten den Film bei der Montage am Schneidetisch noch einmal neu.

Da sich jedes Bild an einen Blickpunkt, an eine Intention anlehnt, schafft jeder Einstellungswechsel einen neuen Bezug, einen nicht mehr isolierten, sondern kontextuellen Sinn (‹nun›, ‹dennoch›, ‹als plötzlich›, ‹aber andernorts› etc.), der das Verständnis des Zuschauers leiten soll.

3 Die Töne aus dem Off können vorhersehbare oder unerwartete Gegenstände evokieren, präzise oder unpräzise Bilder auslösen, die sich in den nächsten Einstellungen materialisieren – oder auch nicht.



1 In der Montage ersetzt der zeitliche Kontext den räumlichen. Das vorangehende und das nachfolgende Element rahmen das dazwischen liegende und bestimmen seine Bedeutung.

Genauso ist es mit den Sequenzen und Akten, insofern die Erinnerungen und Emotionen des Zuschauers in einer Art innerer Montage den Film konstruieren.

Am Schneidetisch steht man multiplen, oft widersprüchlichen Zeitlichkeiten gegenüber, denn die verschiedenen Figuren und Situationen müssen auf einer einzigen Bildspur untergebracht werden, wobei die Qualität der Inferenzen (dass man alles versteht...) und Erinnerungen (dass man noch weiß...) zu bedenken sind. Deswegen gibt es Spannungen zwischen folgenden Zeit(lichkeit)en:

- Die *referenzielle* (Uhr-)Zeit: die exakte Dauer einer Handlung, die vom eigenen Verlauf rhythmisiert wird und innerhalb einer Einstellung nicht komprimierbar ist.
- Die zur *perzeptiven* Aneignung nötige Zeit: das Erkennen von Nahem und Fernem, von Vorder- und Hintergrund, das Erfassen der relativen Geschwindigkeiten. Jede Einstellung verfügt über einen Kontext und zwei Übergänge (*in, out*), die ihr Sinn verleihen und zu ihrer Verständlichkeit beitragen.
- Die *inferenzielle* Zeit: Jede Geste, jede Bewegung entwirft eine nahe oder ferne Zukunft und bezieht sich auf eine jüngere oder ältere Vergangenheit. Signifikante Momente sind genauso an Spuren und Erinnerungen gebunden wie an Fortsetzungen und Postulierungen. Aber nicht alle Momente haben die gleiche Tragweite, die gleiche Kraft, etwas in Gang zu bringen oder auf etwas zurück zu blicken.
- Die *dramaturgische* Zeit: Welcher Figur soll man prioritär folgen, wem sich anschließen? Welche Intentionen und Reaktionen der Protagonisten zeigt man? Wie organisiert man das relative Wissen (weniger oder mehr zu wissen als...), von dem Erwartungshorizonte und Überraschungen abhängen?
- Die *affektive* Zeit: Wann empfiehlt sich eine subjektive Kamera? Welche Dauer gesteht man den Emotionen der Figuren zu (*subito forte, poco a poco crescendo, subito piano*)? Die Emotion des Zuschauers schreibt sich in eine kumulative Zeit (ähnlich der der Musik) ein und nicht in die reale Zeit einer hormonellen Entladung.*

* [Anm.d.Ü.:] Bailblé spielt hier auf Dolf Zillmanns sogenannte «Excitation-Transfer»-Theorie an. Deren Ausgangspunkt ist die Tatsache, dass neuronal vermittelte Kognitionen ein deutlich schnelleres Verlaufsmuster aufweisen als hormonell vermittelte sympathische Erregungen. Das somatisch-hormonelle System ist träger als das flexiblere kognitive System, das auch die Verarbeitung emotionaler Medieninhalte leistet. Der Filmzuschauer muss die größere emotionale Input-Frequenz kognitiv verarbeiten, während auf hormoneller Ebene die Erregungsrückstände kumulieren.

- Die *Reflexions-Zeit*: Der Zuschauer muss an manchen (kürzeren oder längeren) Schlüsselmomenten der Situations- oder Ideen-Entwicklung durchatmen und nachdenken können.

So schichten und entwickeln sich die Dinge im Laufe der Vorführung: Sie münden in einer von Gefühlen und Erkenntnissen überlagerten Konstruktion, einem Zusammenspiel, das die Spuren und Andeutungen variiert, einem unaufhörlichen Neuentfalten von Erwartetem und Unerwartetem, von Projektionen und Introjektionen. Die Montage kann sich also nicht auf die rein deskriptive, mechanische Wiedergabe von Aktionen beschränken, sondern muss folgende Faktoren in Betracht ziehen:

a) Die Kombination von Inferenzen:

- Inferenzen aufgrund von Bewegung (Augen, Gesicht, Körper, Landschaft: also vier verschiedene Geschwindigkeiten);
- Inferenzen aufgrund von Sprache (impliziter oder expliziter Text, stimmliche Gesten, vom Text suggerierte Bilder);
- Inferenzen aufgrund von Geräuschen (im On oder im Off), die plötzlich oder wiederholt auftauchen können oder deren Quelle unsichtbar bleibt;
- Inferenzen aufgrund von Musik (als emotionale Begleitung, als Appell an das körperlich Imaginäre);
- Inferenzen aufgrund der Montage (der Vorher-/Nachher-Kontext ersetzt den räumlichen Kontext).

b) Die Erinnerungen und ihre Dauer:

- Erinnerung an bildliche und klangliche Sinneseindrücke: 0,5 beziehungsweise 2,5 Sekunden;
- Erinnerung an die Einstellung und ihre internen Gliederungen: variable Dauer, normalerweise etwa 10 Sekunden;
- Erinnerung an die einzelnen (trotz der Ellipsen durch den Ton zu einer Einheit verschmolzenen) Sequenzen: meist mehrere Minuten;
- Erinnerung an ganze Akte (gekennzeichnet durch eine Umschichtung des Erwartungshorizonts): etwa 30 Minuten;

So entsteht nach Zillmann ein künstlich erhöhtes Erregungsniveau, das die intensive Wirkung filmisch vermittelter Emotionen erklären könnte. Vgl. Peter Ohler und Gerhild Nieding: *Medienpsychologie*. In: *Psychologie. Eine Einführung in ihre Grundlagen und Anwendungsfelder*. Hg. v. Astrid Schütz, Herbert Selg & Stefan Lautenbacher. Stuttgart: Kohlhammer 2005, S. 453–472, hier S. 466f.

- Erinnerung an den Film insgesamt: Sie währt von Anfang bis Ende der Geschichte;
- Erinnerung an Filme desselben Autors oder an Werke, die in direkter oder indirekter Beziehung dazu stehen, allgemeine Erinnerungen an schon gesehene Filme (kulturelle Prägung der Zuschauer): Diese hält Jahrzehnte an;
- Erinnerung an vom Zuschauer Erlebtes, so wie es persönlichkeitspezifisch – mehr oder weniger lückenhaft und verzerrt – im Gedächtnis gespeichert ist. Vermittelt über die Figuren und Situationen reaktiviert der Film vergessenes Wissen, Ideologien, Wünsche und Gefühle.⁴

Die Mise-en-scène bearbeitet den Zuschauer und lässt ihn arbeiten – bewusst oder unbewusst. Die Erinnerungen aus dem eigenen Leben, in der realen Zeit begegnen den filmischen Erinnerungen in der kondensierten und neu komponierten Zeit der Erzählung. Um dieses Aufeinandertreffen der Projektionen auf der Leinwand mit jenen des Publikums genauer zu charakterisieren, erscheint es mir nötig, das bewegte Bild, das zusammen mit der Montage das Kinodispositiv erst mit Leben erfüllt hat, auf neue Weise zu problematisieren.

Das bewegte Bild

Zum besseren Verständnis des bewegten fotografischen Bildes gehe ich davon aus, dass es vier Prinzipien der Reduktion in sich birgt, die den Raum betreffen, und gleichzeitig ein Prinzip der Expansion, das die Zeit betrifft.

Um mit der Reduktion zu beginnen: Es ist festzuhalten, dass sich das Bild auf eine einzige Ansicht beschränkt, wobei der Vordergrund den Hintergrund überlagert, die Größe der Objekte mit der Distanz abnimmt und Entferntes undeutlich wird. Es nützt nichts, den Kopf zu wenden oder sich zu bewegen: Das Leinwandbild verändert sich nicht. Es handelt sich also um eine reduzierte Ansicht, die sozusagen geronnen ist, was jedoch durch die Montage mittels Vervielfachung der Blickpunkte, der Achsen und Einstellungsgrößen kompensiert und sublimiert wird. Dennoch bleibt jedes Bild unvollständig, weil es kadriert und durch die visuelle Pyramide, die sich vom Blickpunkt bis ins Un-

⁴ Der Film erreicht auch das azeitliche Gedächtnis des Unbewussten: Manche Bilder und Töne verfolgen die Imagination (oder die Träume) noch lange nach der Vorführung.



2 Die Landschaft (3D) erzeugt in der koplanaren fotografischen Abbildung einen ambivalenten räumlichen Eindruck.

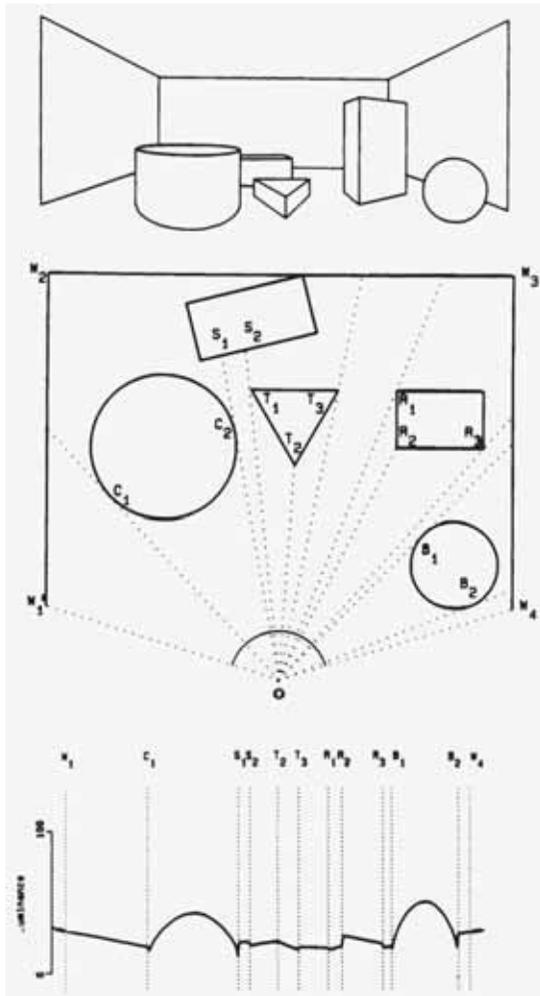
endliche erstreckt, begrenzt ist. Dies bewirkt, dass das Off – der Kontext – dem Vorstellungsvermögen des Betrachters überlassen bleibt.

Des Weiteren verdichtet sich die Ansicht (der Blickwinkel) zu der bewussten Fensterperspektive in dem Sinne, dass die Leinwand nur ein künstliches – koplanares – Bild bietet, d.h. ein Bild, das weder Auswölbung noch echte Tiefe aufweist. So entsteht ein Regime des allgemeinen Nebeneinanders, das durch die große Tiefenschärfe noch verstärkt wird: Figuren und Hintergründe verbinden sich auf der Bildfläche und gehen Beziehungen und Nachbarschaften ein, wie wir sie in der Wirklichkeit nicht wahrnehmen.

Wie sich leicht zeigen lässt, verfügt das direkte Sehen normalerweise nur über eine einzige Schärfezone, die sich am Punkt der Überschneidung der beiden okularen Achsen befindet. Außerhalb dieser Fusions Ebene wird die binokulare Verdoppelung durch einen kortikalen Mechanismus unterdrückt, mit anderen Worten: Die Realität wird außerhalb der fokussierten Zone nur sehr schwach wahrgenommen – dies im Unterschied zu einem Bild von großer Tiefenschärfe.

Aufgrund der Unmöglichkeit, die binokulare Konvergenz bei der selektiven Analyse der verschiedenen Tiefenschichten ins Spiel zu bringen, werden schließlich die Wahrnehmungskonstanten hinsichtlich der Größe und der Form außer Kraft gesetzt. Daraus ergibt sich dann die besondere Bedeutung der Einstellungsgrößen und der Ausrichtung der Gegenstände für den Blick der Kamera. Deshalb ist bei der Filmaufnahme ein Gespür für den Blickwinkel erforderlich, damit

3 Oben eine perspektivische Darstellung; in der Mitte dieselbe Anordnung, nunmehr von oben und in Hinblick auf den Betrachter gesehen; unten schließlich die Darstellung des dabei vom Auge wahrgenommenen Helligkeitsprofils.



die Elemente, die man zeigen will, auch im koplanaren Rechteck hervortreten. Das Bild erfordert, da es «mono» – und darüber hinaus auf einer zweidimensionalen Fläche – reproduziert wird, ein perspektivisches Arrangement, eine intensiv durchdachte Komposition.

David Marrs Theorie des Sehens (1982) kann uns helfen, das filmische Bild, das Erbe des Quattrocento, besser zu verstehen.* Nach

* [Anm.d.Ü.:] David Bordwell bietet in «A Case for Cognitivism» (in: *Iris*, 9, 1989,

dieser Theorie «scant» das Licht das Objektfeld und wird nach allen Seiten reflektiert. Die Pupille selektiert daraus ein einfallendes Strahlenbündel – einen Aspekt unter vielen –, das in ein zweidimensionales Bild auf den retinalen Hintergrund projiziert wird. Die dem Licht entnommene Information generiert Textur, Kontur, Farbe, Position, Bewegung und Bewegungsbahn, wenn sie durch den visuellen Kortex analysiert wird. Eine erste Formgebung beginnt jedoch ab der Retina (als Element des Gehirns). Dies ist nach David Marr das 2D-Stadium des Sehens. In diesem Stadium «sieht» die retinale Oberfläche lediglich unterschiedlich helle Bereiche und Lichtwechsel und setzt sie in bestimmte visuelle Eigenschaften und trennende Konturen um.⁵ Durch den optischen Nerv werden diese – retinotopischen – Daten zum visuellen Kortex geleitet.

Mehrere perzeptive Module, die stark verkapselt und deswegen der Kognition schwer zugänglich sind, sorgen dafür, dass die elementaren Eindrücke (die Qualia des Sehens) wie Farbe, Bewegung, Wölbung, Textur oder Schatten sich einstellen, als würden sie der Realität zugehören, die den Betrachter umgibt und deren Zentrum er bildet: mit David Marr zu sprechen ist dies eine in 2,5 D(imensionen) wahrgenommene.

Das 2,5D ist die notwendige, zwischengeschaltete Etappe zum 3D, die gekennzeichnet ist von einer Vorherrschaft des Visuellen gegenüber dem Realen, einer Prävalenz des Blickpunkts. In diesem Stadium sieht man nur eine Facette der Realität, ein Unterensemble des Sichtbaren, das zwar von zwei Augen und zweifellos auch konvex/konkav wahrgenommen wird, aber eben nur *eine* aus einem bestimmten Winkel gesehene, egozentrierte Ansicht darstellt. Man erkennt die Vorderseiten, nicht aber die Rückseiten, manches ist sichtbar, anderes nicht: Zahlreiche Objekte verdecken einander.

Genau in diesem Stadium arbeitet das Filmbild. Niemandem würde es in den Sinn kommen, sich zur Seite zu neigen oder sich zu bewegen, um zu sehen, was die vordere Bildebene verdeckt: Das Bild ist auf die momentane Ansicht reduziert und erstarrt (wie in der Etappe des 2,5D) und sogar zum koplanaren 2D kondensiert. Jede Einstellung ist daher – im wörtlichen wie im übertragenen Sinn – durch den Blickwinkel der Aufnahme bestimmt: Das filmische Bild kann dem

S. 11–39, hier S. 24f) eine einführende Skizze von David Marrs Theorie; eine ausführlichere Auseinandersetzung liefert Steven Pinker in «Visual Cognition: An Introduction» (in: *Cognition*, 18, 1984, S. 1–63, hier S. 14ff).

5 Die Spur dieser retinalen Arbeit erscheint übrigens in der Strichzeichnung oder in der Schrift (die Minimalinformation: die schwarze Tintenlinie auf weißem Papier).

intentionalen und konstruierten Charakter dieses Blicks so wenig entkommen wie der mit der Projektion gegebenen, automatischen Zentrierung des Betrachters.

Daher die Bedeutung, die der letzten, 3D genannten Phase zukommt, in der sich der reale Beobachter (aber nicht der Filmzuschauer) wieder vom Bildeindruck löst. Das ist der Moment, an dem sich der Blick als solcher vom visuellen Feld zurückzieht, um eine vom retinalen 2D-Bild und der Aspektualität des 2,5D befreite räumliche Realität zu schaffen. Nun scheint sich das Volumen der Gegenstände ganz von allein zu ergeben, die Lebewesen und Dinge sind an ihrem Platz, dezentriert, in ihren exakten Dimensionen. Wir sind «im realen Raum», die Qualia des Sehens, welche die vor uns erscheinende Welt sozusagen «bekleiden», haben wir vergessen.

Das Bild des Blickfeldes wurde also automatisch an seinen Ausgangspunkt zurückprojiziert, um mit der geometrischen Realität der Dinge zu verschmelzen und die Volumina komplett zu umhüllen (selbst die verdeckten Partien). In 3D finden die dreidimensionalen Gegenstände ihre realen Maße, ihren Ort und ihre Farbe, also eine Ding-Existenz und keine Bild-Existenz, mit anderen Worten: Der Raum erscheint in seiner wahren Dimension und mit den korrekten Abständen zwischen den Gegenständen.

Das fotografische Bild ist also nur teilweise objektiv, da es unvollständig und reduziert ist (was Räumlichkeit und Helligkeit angeht): Wenn ein Grenzbeamter einen Pass verlangt, betrachtet er das Foto eines unbekanntes Gesichts. Er muss das Original (in 3D) mit dem Miniaturfoto (2,5D auf 2D komprimiert) vergleichen, um die Ähnlichkeit zu überprüfen und die Identität festzustellen. Er vergleicht zwei Bilder: Das eine, vom Gehirn erarbeitete, wird «real» genannt, das andere als «Foto» bezeichnet, als «ikonisches Zeichen» oder gar «Analogon».

Doch die Bedeutung des Bildes wird zugleich von der subjektiven Interpretation oder dem subjektiven Kontext des Betrachters geprägt, denn sie hängt von dessen vorherigem Wissen, von Verkennungen oder auch Unkenntnissen ab. Sie ist sogar doppelt subjektiv: einmal bereits bei der Aufnahme, die der Fotograf mit einer bestimmten Absicht und aus einem räumlichen wie zeitlichen Kontext heraus macht. Er wird versuchen, sich das Resultat – die Schnitte, die Kadrierung, die koplanare Ansicht – schon vorzustellen, bevor er den Auslöser betätigt. Dann ein weiteres Mal bei der Rekonstruktion durch den Betrachter, der dem Bild Sinn verleihen muss, ohne den Kontext der Aufnahme zu kennen. Er verlässt sich dabei einerseits auf allgemeine Gegebenheiten und indirektes Wissen, andererseits auf den koplanaren Verbund

der Figuren und Hintergründe, die perspektivische Hierarchisierung der vorderen und hinteren Bildebenen und die auf 2,5D geronnene Ansicht. Im Kino stiftet die Montage den Kontext: als unaufhörliche Akkumulation von Spuren des Gewesenen und Andeutungen des Kommenden.

Im Zusammenhang mit dem bewegten Bild habe ich bereits auf ein Prinzip der Expansion hingewiesen. Dieses Prinzip, das man als «inferenziell» bezeichnen kann, erlaubt es dem Zuschauer – je nach seinen intellektuellen Fähigkeiten und seiner Aufmerksamkeitsspanne –, die Ursachen eines signifikanten Vorgangs, einer prägnanten Geste oder einer Bewegung zu begreifen oder deren Konsequenzen zu antizipieren. Aufgrund der Veränderung im Bild und dank der damit verbundenen simplen kognitiven Erregung versucht der Zuschauer, den beobachteten Vorgang in eine Kausalkette einzuordnen, indem er ihn zum Vorhergegangenen oder Folgenden in Beziehung setzt.

Deshalb besteht keine Notwendigkeit, eine filmische Handlung in all ihren Phasen und in ihrer vollen Dauer zu zeigen; es reicht aus, die entsprechenden Momente zu präsentieren, damit der Zuschauer sich das Kommende vorstellen oder das zuvor Geschehene ableiten kann. Darin verbirgt sich ein Prinzip narrativer Ökonomie, das Ellipsen und insbesondere Parallelmontagen ermöglicht, bei denen man mehreren Figuren gleichzeitig folgen kann, ohne den Faden zu verlieren, indem die Erzählperspektive alterniert, die Intentionen oder Reaktionen aller Beteiligten abwechselnd zur Geltung kommen oder die Ereignisse sich an bestimmten Orten überschneiden. Die Augen, das Gesicht (die personale Achse) sowie die Körper und Bewegungen in der Landschaft (die situative Achse) sind also ausdrucksstark genug, um es dem Zuschauer zu ermöglichen, ihre Dynamik über das räumlich und zeitlich Gegebene hinaus – ins Off bzw. in das Davor oder Danach – weiterzudenken. Das Verstehen übersteigt die unmittelbare Wahrnehmung.

Schon bei den Dreharbeiten gilt es deshalb, ein Gespür für die zeitliche Auflösung einer Szene zu entwickeln: Es geht darum, passende Momente einzufangen oder zu konstruieren, die der Einstellung ihre Entwicklung und ihren Rhythmus verleihen und Anschlussmöglichkeiten für die folgende Einstellung bieten. Bei der Montage ist diese Fähigkeit ein zweites Mal gefragt: Nun müssen Dauer und Platzierung der Einstellungen bestimmt und die Momente ausgewählt werden, die sich sukzessiv zu einer Art erweiterter Gegenwart zusammenfügen sollen. Untrennbar mit dem Blickpunkt verbunden, der die durch die Perspektive vereinten und konfigurierten Elemente verstehbar macht, wird sich das durch die Einstellung provisorisch geöffnete Zeitfenster

nicht schließen, bevor nicht vorherige Elemente weitergeführt oder folgende vorbereitet worden sind.

Lange habe ich mich gefragt, wieso wir so viel schneller begreifen als wahrnehmen können. Würde man die gesamte Handlung in ihrer Chronologie und Entwicklung verfolgen, so würde die Wahrnehmung dabei in ihrem Verlauf mehrfach vom Verstehen überholt. Der Zuschauer würde ungeduldig scharren und stampfen und eine raschere Entwicklung fordern. Ohne Frage sind die von der Bewegung (der Augen, der Hände und des übrigen Köpers) ausgelösten Inferenzen schneller als die Bewegung selbst. Aber die einfache Tatsache, dass diese Inferenzen vorbewusst bleiben, dass sie wenig Aufmerksamkeit beanspruchen und gleichzeitig ein so schnelles und verlässliches Ergebnis liefern, kann einen immer wieder in Erstaunen versetzen.

Eine Theorie, die 1997 von einer Forschergruppe der Universität Parma entwickelt wurde, könnte auf diese Frage eine überzeugende Antwort liefern. Die Befunde dieser Gruppe besagen, dass einige Zellen des Gehirns, die sogenannten «Spiegelneuronen», in gleicher Weise aktiviert werden, wenn man eine Handlung *beobachtet* und wenn man sie *selbst ausführt*.

Peter sieht, wie Anna nach einer Blume greift. Er weiß sofort, dass sie die Blume pflücken möchte – und mehr. Denn sie lächelt ihm zu. Er merkt: Sie möchte ihm die Blüte schenken. Wie kann es sein, dass Peter Annas Absicht fast unmittelbar erahnt, kaum dass die kurze Handlung begonnen hat? (Rizzolatti/Fogassi/Gallese 2007, 49).

Noch vor zehn Jahren hätte man das Verstehen einer solchen Handlung und deren Absicht einer unmittelbaren logischen Schlussfolgerung zugeschrieben, wobei die wahrgenommene Aktion mit einer ähnlichen, schon im Gedächtnis gespeicherten verglichen würde. Heute ist die Erklärung sehr viel einfacher: Dieselben Neuronen des Parietal- und prämotorischen Kortex werden aktiviert, wenn man eine Handlung ausführt und wenn man sie beobachtet: «Es wirkte, als würden diese Zellen den beobachteten Verhaltensakt direkt, unmittelbar widerspiegeln. Deswegen nannten wir sie «Spiegelneuronen» oder «Spiegelzellen» (ibid., 49).

Wenn wir – wie so oft – handeln, ohne nachzudenken, oder wenn wir Gesten, die wir kennen, unmittelbar begreifen, liegt das daran, dass wir gerade nicht bewusst nachdenken müssen. Seit langem weiß man, dass es ausreicht, eine Situation zu beobachten, um sie zu verstehen. Das Spiegelneuronensystem, das als Gedächtnis des eigenen Handelns

fungiert und zugleich alles, was man im Begriff ist zu sehen und zu hören, interpretiert, bestätigt diese Intuition. Doch die Theorie besagt noch mehr:

Genauer betrachtet ist eine Bewegungsabfolge von ihrem Ziel bestimmt, von ihrer Intention. Es ist etwas anderes, ob ich beabsichtige, an der Blüte zu riechen, oder ob ich die Blume pflücke, um sie jemandem zu überreichen (ibid., 52).

Die wichtigste Entdeckung der drei italienischen Forscher bestand darin, im bildgebenden Verfahren der funktionalen Kernspintomographie zu zeigen, dass die Spiegelneuronen die Intentionen des Anderen «verstehen», indem sie ähnliche Handlungen, die jedoch unterschiedliche Ziele haben, voneinander unterscheiden. So löst zum Beispiel die Handlung «nach einer Tasse greifen, um zu trinken» nicht die gleiche neuronale Aktivität aus wie «nach einer Tasse greifen, um den Tisch abzuräumen».

Die Spiegelneuronen reagieren mit anderen Worten vor allem auf die intentionalen Komponenten, indem sie eine Beziehung zum Handlungskontext herstellen. Diese selektive Reaktion ermöglicht es, das Verhalten Anderer unmittelbar zu verstehen, ohne auf ein semiologisches Repertoire von Blicken, Gesten oder Bewegungen zu rekurrieren; es genügt, sich auf das körperliche Gedächtnis und die intentionalen Schemata zu stützen, die im Neuronennetzwerk eingraviert und kondensiert sind.

Diese Reaktion erlaubt auch den Zugang zu den Gefühlen der Filmfiguren, denn der Spiegelmechanismus stimuliert ein Mitempfinden ihres emotionalen Zustands, unabhängig davon, ob er sich auf den in Großaufnahme gezeigten Gesichtern oder in einer Halbtotalen an den Bewegungen und Haltungen der Körper manifestiert. Die solcherart mitgefühlten Emotionen und Intentionen bereiten den Zuschauer direkt auf die subjektive Kamera vor.

Durch die Empathie mit den Figuren nehmen wir an deren Intentionen und Reaktionen teil, wobei die Geschwindigkeit unseres Verstehens dem Tempo der narrativen Zeit entspricht und letztlich von der Zeit der Aktivierung des Neuronennetzwerks abhängt (der Übertragungsdauer vom visuellen zum prämotorischen Kortex und zurück zum beobachteten Ereignis). Dabei ist verblüffend, wie verkürzt und dicht die prämotorische Repräsentation ist und fein und bestimmt dennoch.

Die prämotorischen Bilder werden vom Handelnden nur «zwischenwahrgenommen», denn der sensomotorische Dialog mit der Um-

gebung darf nicht abreißen. Tatsächlich wäre es sehr riskant, wenn die antizipatorische Vorstellung einer Handlung die unmittelbare Wahrnehmung in großer Intensität überlagern würde. Übrigens braucht das Gehirn oft nur Sekundenbruchteile, um die Konsequenzen einer Geste abzuschätzen und zu einer Entscheidung zu kommen.

Zweifellost ist es – wie im Kino – viel einfacher, die Handlungen Anderer vorzusehen, indem man sich in ihre Intentionen einfühlt, als sich in klarer Weise die Folgen eigener Handlungen vorzustellen. Treffen die durch die Montage suggerierten Absichten der Protagonisten auf Hindernisse oder werden von Unvorhergesehenem durchkreuzt, so spielen auf jeden Fall mentale Bilder (möglicher) motorischer Prozesse eine Rolle. Die *prämotorische* Bilderwelt nehmen wir sicherlich immer nur wie ein Gespinnst wahr, sie bleibt vorbewusst. Erst wenn sich eine Schwierigkeit oder ein Problem ergibt, schaltet sich das Bewusstsein ein. In einem solchen Moment wird das Gespinnst gleichsam von innen erhellt, und das Bild gleitet vom dispositionalen Gedächtnis (das der Verkettung von Phasen oder Gesten dient) zum repräsentierenden Gedächtnis (das Situationen und Resultate speichert).⁶

Alles in allem lässt sich sagen, dass die prämotorische Bilderwelt – so fragil und fein sie auch sein mag – in Kombination mit dem flüchtigen Bild des Ziels einer Handlung oder auch mit dem detaillierten mentalen Bild des Hindernisses die Zukunft (das Bild der Intention) in die Gegenwart (das wahrgenommene Bild) hereinholt.⁷ Auch wenn sie von bedeutsamen Momenten, von den durch die *Mise-en-scène* gewählten Blickwinkeln oder auch von der Qualität des Schauspiels abhängt, vollzieht sich die praktisch augenblickliche Aktivierung gespeicherter intentionaler Muster viel schneller und ergiebiger als der objektive Lauf der Dinge. Dies zumindest, wenn das Kino an das Imaginäre appelliert, das heißt an die Erfahrung des einzelnen Zuschauers, der letztlich der Interpret des bewegten Bildes ist.

Versunken in eine Szenenfolge, die er ohne Unterlass durch Introjektion und Projektion konstruiert, erlebt der Zuschauer – in einem

6 Die unaufhörliche Vermischung von submotorischen Erlebnissen und antizipierten Repräsentationen konstituiert in meinen Augen eines der Spezifika des Actionfilms. Der Zuschauer erschauert, stampft mit den Füßen, erstarrt und entspannt sich. Der Körper identifiziert sich nicht nur mit den Emotionen und Intentionen, sondern mittels der skizzierten Motorik auch mit den Bemühungen und Befürchtungen der Figuren.

7 Diesen Prozess versuche ich in meiner Dissertation (Bailblé 1999, 193ff) – aus der ich hier zitiere – ausführlicher zu erläutern. Ich beziehe mich dort vor allem auf die Arbeiten von Jacques Paillard (1994).

durch die Narration intensivierten Zustand der Empathie – die Wünsche und Projekte der Figuren, die stellvertretend für ihn handeln und mit denen er sich identifiziert. Aber die Tatsache, dass hier im Modus des «als ob» innerhalb eines fiktionalen, meist von typisierten Figuren getragenen Dispositivs agiert wird, lässt die Frage offen, inwieweit der Zuschauer die imaginär und emotional miterlebten Handlungen noch als eigene empfindet, wenn er in seine Alltagsrealität zurückgekehrt ist.

Aus dem Französischen von Guido Kirsten

Literatur

- Alberti, Leon Battista (2002) *Della Pittura. Über die Malkunst* [lat. 1435]. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Baillblé, Claude (1999) *La perception et l'attention modifiées par le dispositif cinéma*. Diss. Paris: Université de Paris VIII.
- Burch, Noël (1990) *Life to those Shadows*. Berkeley/Los Angeles: University of California Press.
- Marr, David (1982) *Vision*. San Francisco: Freeman and Company.
- Paillard, Jacques (1994) L'intégration sensori-motrice et idéo-motrice. In: *Traité de psychologie expérimentale I*. Hg. v. Marc Richelle, Jean Requin & Michèle Robert. Paris: PUF, S. 925–955.
- Rizzolatti, Giacomo / Fogassi, Leonardo / Gallese, Vittorio (2007) Spiegel im Gehirn. In: *Spektrum der Wissenschaft*, 3, S. 49–55. [Zuerst englisch als: Mirrors in the Mind. In: *Scientific American* 295,5, 2006, S. 30–37.]