

# **Geographie, Raum und Geographische Informationssysteme**

**Notwendigkeit und Entwicklung eines verallgemeinerten  
Algorithmus zur Verschneidung von Flächen**

Dissertation zu Erlangung der Würde eines Dr. rer. pol.  
an der Universität Bremen

Vorgelegt von Dipl. Geogr. Nils Mevenkamp  
im Juni 1999

1. Gutachter: Prof. Dr. G. Bahrenberg

2. Gutachter: Prof. Dr. F. Nake

Tag des Promotionskolloquiums: 15. Juli 1999

Ich versichere, daß ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe angefertigt habe, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und alle den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Nils Mevenkamp  
Fesenfeld 123  
28203 Bremen

# Danksagung

Die Ideen, die in dieser Arbeit vorgetragen werden, gehen zu einem großen Teil auf Diskussionen mit FreundInnen und KollegInnen zurück. Herrn Prof. Dr. Gerhard Bahrenberg danke ich für die gleichermaßen fachlich kompetente und persönliche Unterstützung und Herrn Prof. Dr. Frieder Nake für die Annahme der Arbeit und insbesondere für wertvolle Hinweise bezüglich des methodischen Teils. Ein ganz besonderes Dankeschön geht an Kerstin Albers, die mein liebevoller Rückhalt und meine schärfste Kritikerin war. Janine Geisler, Frank R. Liebich (der den wunderbaren Begriff des ‘Integrations-Systems’ geprägt hat) und Dieter Heilbronn haben weitere Korrekturen übernommen, Dr. Ralf Zarsteck hat mir die Kartengrundlage und die Daten für das Neustädter Beispiel zur Verfügung gestellt. Viele sind am Gelingen dieser Arbeit beteiligt gewesen, und nur einige von ihnen können an dieser Stelle ausdrücklich genannt werden; es sei jedoch allen gedankt.

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	GIS als Geographisches Informations-System . . . . .	1
1.2	GIS als Geographischer Informations-Szientismus . . . . .	4
1.2.1	Die Charakteristik erdraumbezogener Daten . . . . .	5
1.2.2	Spatial Analysis . . . . .	5
1.2.3	Geoinformatik . . . . .	7
1.2.4	Spatial Information Theory . . . . .	8
1.3	GIS als Geographisches Integrations-System . . . . .	9
1.4	GIS als Geometrisches Imaginations-System . . . . .	14
<b>2</b>	<b>Zur Funktionsweise geometrischer Abstraktionen</b>	<b>18</b>
2.1	Navigation . . . . .	19
2.1.1	Navigation aufgrund physischer Wahrnehmung . . . . .	19
2.1.2	Navigation aufgrund symbolischer Wahrnehmung . . . . .	19
2.1.3	Kognitive Karten . . . . .	20
2.1.3.1	Kognitive Karten und Geographie . . . . .	23
2.2	Gedächtnis . . . . .	24
2.2.1	Kurz- und Langzeitspeicher . . . . .	25
2.2.1.1	Das modal model . . . . .	25
2.2.1.2	Zweitaufgabentechnik . . . . .	26
2.2.2	Arbeitsgedächtnis . . . . .	27
2.2.2.1	Die artikulatorische Schleife . . . . .	28
2.2.2.2	Der räumlich-visuelle Notizblock . . . . .	28
2.2.3	Bilder und Vorstellungen . . . . .	29
2.2.4	Kognitive Karten und Bilder . . . . .	31
2.3	Deduktion . . . . .	33
2.3.1	Beweistheoretische Erklärungen . . . . .	33
2.3.2	Mentale Modelle . . . . .	34
2.3.2.1	Syllogistisches Schließen und der figurale Effekt . . . . .	34
2.3.2.2	Das Drei-Phasen-Modell des Schlußfolgerns . . . . .	35
2.3.2.3	Evidenz mentaler Modelle . . . . .	38
2.3.2.4	Mysteriöse Logik . . . . .	40
2.3.2.5	Mentale Räume . . . . .	41
2.3.3	Kognitive Semantik . . . . .	43
2.3.3.1	Kinästhetische Erfahrungen und Bild-Schemata . . . . .	44
2.3.3.2	Logische Diagramme . . . . .	45
2.4	Diagrammatisches Denken . . . . .	46

<b>3</b>	<b>Die Bestimmung planarer Subdivisionen</b>	<b>50</b>
3.1	Problemstellung und Begriffsdefinitionen . . . . .	50
3.2	Die Verschneidung zweier Flächenkarten . . . . .	54
3.2.1	Der Scan-Line-Algorithmus . . . . .	55
3.2.2	Die doppelt-verkettete Kantenliste . . . . .	61
3.2.3	Berechnung der Schnittflächen aus zwei planaren Subdivisionen . . . . .	64
3.2.3.1	Aktualisierung der Knoten- und Kantentabelle . . . . .	65
3.2.3.2	Aktualisierung der Flächentabelle . . . . .	68
3.2.3.3	Aktualisierung der Flächenlabel . . . . .	70
3.2.3.4	Ein einfacher Verschneidealgorithmus . . . . .	72
3.2.4	Berechnung von Sekundärdaten . . . . .	74
3.3	Die Verschneidung von $m$ Flächenkarten . . . . .	75
3.3.1	Problemstellung . . . . .	75
3.3.2	Bestimmung der Flächenlabel . . . . .	77
3.3.3	Bestimmung der umgebenden Flächen . . . . .	78
3.3.4	Zeitaufwand . . . . .	82
3.4	Die Handhabung sich überlappender Kanten . . . . .	83
3.4.1	Bestimmung sich überlappender Kantenpaare . . . . .	83
3.4.2	Aktualisierung der Flächenlabel . . . . .	84
3.4.3	Eine Typisierung von Überlappungsproblemen . . . . .	85
3.4.4	Die algorithmische Bearbeitung von Überlappungsproblemen . . . . .	86
3.4.5	Zeitaufwand . . . . .	88
3.5	Umgang mit unpräziser Arithmetik . . . . .	88
3.5.1	Primitive arithmetische Operationen . . . . .	92
3.5.1.1	Berechnung des Schnittpunkts zweier Liniensegmente . . . . .	92
3.5.1.2	Berechnung der relativen Lage eines Punktes . . . . .	94
3.5.2	Repräsentationen mit fester Präzision . . . . .	96
3.5.2.1	Modelle und Repräsentationen . . . . .	97
3.5.2.2	Symbolisches Schlußfolgern mit unpräziser Arithmetik . . . . .	98
3.5.2.3	Schnittpunktrepräsentationen und Präzision . . . . .	101
3.5.3	Verschneidung mit topologisch konsistenter Verzerrung . . . . .	103
3.5.3.1	Lokale Verletzungen der Invariante . . . . .	104
3.5.3.2	Globale Verletzungen der Invariante . . . . .	111
3.5.4	Handhabung von (Fast-)Überlappungen . . . . .	112
3.5.5	Aktualisierung der Flächenlabel . . . . .	112
3.5.6	Ein robuster Verschneidealgorithmus . . . . .	114
3.5.7	Zeitaufwand . . . . .	125
<b>4</b>	<b>Anwendungsbeispiele</b>	<b>127</b>
4.1	Erreichbarkeit sozialer Infrastruktur . . . . .	127
4.1.1	Durchmischung und Konzentration sozialer Infrastruktur . . . . .	128
4.1.2	Relativer Versorgungsgrad . . . . .	129
4.2	Visuelle Regressionsanalyse . . . . .	136
<b>5</b>	<b>Zusammenfassung</b>	<b>140</b>
	<b>Literaturverzeichnis</b>	<b>143</b>

# Abbildungsverzeichnis

2.1	Euler-Kreise . . . . .	46
2.2	Lakoffs Interpretation klassischer Kategorien als Container . . . . .	47
2.3	Grundkonfigurationen eines Venn-Diagramms . . . . .	49
3.1	Eine nicht verbundene planare Subdivision . . . . .	51
3.2	Eine visuelle Verschneidung . . . . .	52
3.3	Eine geometrische Situation vor und nach der Verschneidung . . . . .	55
3.4	Verschiedene Anordnungen von Liniensegmenten . . . . .	56
3.5	Drei aufeinanderfolgende Eventpunkte eines ‘Scanline-Schwenks’ . . . . .	58
3.6	Ein Segmentschnittproblem und seine Lösung mit Hilfe des Scanline-Verfahrens . . . . .	61
3.7	Eine doppelt-verkettete Kantenliste . . . . .	63
3.8	Die Aktualisierung der doppelt-verketteten Kantenliste $\mathcal{D}$ . . . . .	65
3.9	Die Teilung einer Kante . . . . .	67
3.10	Das ‘Linksabbieger-Prinzip’ beim Aktualisieren der Flächeninformation . . . . .	68
3.11	Die Zuordnung von äußeren und inneren Flächengrenzen . . . . .	70
3.12	Die Bestimmung der Flächenlabel . . . . .	71
3.13	Themabasierte Verschneidungen mit Hilfe Boolescher Junktoren . . . . .	74
3.14	Eine visuelle Verschneidung von vier Karten . . . . .	76
3.15	Die Bestimmung der Flächenlabel für $m > 2$ . . . . .	78
3.16	Das Problem der umgebenden Flächen . . . . .	81
3.17	Das Problem zweier sich überlappender Kanten . . . . .	83
3.18	Die Aktualisierung von überlappenden Kantenpaaren . . . . .	84
3.19	Eine Typisierung von Überlappungsproblemen . . . . .	85
3.20	Verschiedene Vergrößerungen von 327 Zwölfecken . . . . .	90
3.21	Die Bestimmung des Schnittpunktes zweier Liniensegmente . . . . .	93
3.22	Die Bestimmung der relativen Lage eines Punktes . . . . .	94
3.23	Die Konstruktion von Liniensegmenten fester Präzision . . . . .	96
3.24	Mögliche Folgen unsicherer numerischer Daten . . . . .	99
3.25	Modelle und Repräsentationen . . . . .	100
3.26	Modelle und Repräsentationen von Schnittpunkten . . . . .	101
3.27	Die Präzision einer Verschneidung mit verzerrter Topologie . . . . .	102
3.28	Überziehung von Kanten als Folge verzerrter Topologie . . . . .	105
3.29	Überziehung von Knoten als Folge verzerrter Topologie . . . . .	106
3.30	Versetzung von Eventpunkten als Folge verzerrter Topologie . . . . .	111
3.31	Die Bestimmung der Flächenlabel bei konsistent verzerrter Topologie . . . . .	113
4.1	Standorte von sozialen und Bildungseinrichtungen in der Bremer Neustadt . . . . .	130
4.2	Die Bevölkerungsdichte in der Bremer Neustadt . . . . .	131

4.3	Die Flächen der Baublöcke und Einzugsbereiche in der Bremer Neustadt . . . . .	132
4.4	Durchmischungsgrad von sozialen und Bildungseinrichtungen in der Bremer Neustadt	133
4.5	Konzentration von sozialen und Bildungseinrichtungen in der Bremer Neustadt . . . .	134
4.6	Relativer Versorgungsgrad mit sozialen und Bildungseinrichtungen in der Bremer Neustadt . . . . .	135
4.7	Das Streudiagramm der Variablen <i>TERT</i> und <i>FRAU</i> . . . . .	136
4.8	Das Streudiagramm der Variablen <i>TERT</i> und <i>FRAU</i> mit kreisförmigen Datenflächen .	138
4.9	Eine visuelle Regressionsanalyse . . . . .	139



# Kapitel 1

## Einführung

### 1.1 GIS als Geographisches Informations-System

Der Umgang mit Geographischen Informationssystemen (GIS<sup>1</sup>), d.h. Computersystemen „zur Erfassung, Verwaltung und Analyse großer Mengen räumlich verorteter Daten und darauf bezogener thematischer Attribute“ (Dollinger, Strobl 1989a, S. 10), ist heute sowohl in der Geographie als auch in vielen nicht-geographischen Disziplinen selbstverständlich. Die Ausgabe des *GeoDirectory Products and Services Purchasing Guide*<sup>2</sup> für die Jahre 1998-1999 listet 337 verschiedene Anbieter sogenannter ‘geospatial products’ auf, und im September 1998 wird das Investitionsvolumen für die folgenden 12 Monate allein in Deutschland auf 196 Mio. DM geschätzt (GIS-Europe 1998/9, S. 22; vgl. auch Schilcher 1996). Der Erfolg Geographischer Informationssysteme ist unbestreitbar, die Entwicklung ist rasant, und der Markt boomt.

Die Grenze zwischen GIS und anderen raumbezogenen Computertechniken (digitale Fernerkundung, Computerkartographie u.a.) ist fließend, und Definitionsversuche erfolgen in der Regel über technische Kriterien (vgl. z.B. Burrough 1986; Goodchild 1995). Zu GIS gehört eine Datenbank, die sowohl reale Objekte (z.B. Wetterstationen, Flüsse, Ackerflächen) in Form ihrer geometrischen Modelle (Punkte, Linien, Flächen) als auch Attribute dieser Objekte (z.B. Höhenlage, Fließgeschwindigkeit, Vegetationsbedeckung) in großen Mengen speichern kann und die die eindeutige Zuordnung von Geometrie- und Attributdaten gewährleistet. GIS beinhaltet Funktionen für die Darstellung und Ausgabe dieser Daten – in erster Linie in Form von Karten, aber auch tabellarisch oder diagrammatisch – auf Bildschirmen, Druckern, Plottern usw. Dazwischengeschaltet sind sogenannte Analysefunktionen, die die eigentliche Stärke und für viele auch die Einzigartigkeit von GIS ausmachen (z.B. Dangermond 1990). Die mit GIS gebotene Analytik setzt immer auf einem Datenmodell auf, das auf *raumbezogene* Fragestellungen zugeschnitten ist, das also bzgl. der geometrischen Objekte und den ihnen zugeordneten Sachdaten die Beschreibung einer räumlichen Struktur erlaubt, z.B. in Form von räumlichen Relationen wie ‘schneidet sich’, ‘ist benachbart’, ‘ist verbunden mit’, ‘ist in der Nähe von’ u.v.m.<sup>3</sup>

Das erste Computerprogramm, das nach allgemein gängiger Auffassung zu Recht als GIS bezeichnet werden kann, war das *Canada Geographic Information System* (Tomlinson 1990), das in den 1960er Jahren im Auftrag der Kanadischen Regierung zum Zwecke der landesweiten Aufnahme

---

<sup>1</sup>Die Abkürzung ‘GIS’ wird sowohl in der deutschsprachigen als auch in der anglo-amerikanischen Literatur gleichermaßen im Singular und im Plural sowie in verschiedenen Bedeutungen gebraucht. Gemeint sind je nach Kontext einerseits konkrete Hard- und Softwaresysteme und andererseits die diesen Systemen zugrundeliegende konzeptionelle Grundidee. Dieser Sprachgebrauch wird auch hier übernommen.

<sup>2</sup>Adams Business Media, [www.GeoPlace.com](http://www.GeoPlace.com)

<sup>3</sup>In der häufig etwas umgangssprachlichen Ausdrucksweise der GIS-Gemeinde werden die genannten Relationen auch als *topologisch* bezeichnet. Für eine vollständige Auflistung siehe Albrecht 1996, S. 61ff.

der tatsächlichen und potentiellen Landnutzung Kanadas entwickelt wurde. Das gesamte Land wurde mittels zahlreicher Variablen – z.B. tatsächliche Landnutzung, Bodentyp, Potential zur agrarischen Nutzung, Potential als Erholungsraum u.v.m. – im Maßstab 1:50.000 inventarisiert. Da jede einzelne Variable an jedem Punkt in Kanada genau einen Wert aufweist, wurden die Daten in Form sogenannter *Layer*<sup>4</sup> organisiert. Die Datenbasis enthielt somit für jede Variable genau einen Layer, der für jeden Punkt der abgebildeten Fläche (des kanadischen Territoriums) genau einen Wert der entsprechenden Variable vorhält.

Im Rahmen der Volkszählung im Jahr 1970 wurde für das *U.S. Bureau of the Census* ein ähnliches System entwickelt. Auch hier bot das Layer-Konzept einen adäquaten Rahmen für das Datenmanagement, denn jeder Ort (Punkt) in den USA liegt in genau einem Bundesstaat, einem County, einem Census Tract usf. (vgl. Goodchild 1995, S. 37; Tomlinson 1990).

Ausgangspunkt des für die Arbeit mit GIS heute so akzeptierten „layer-cake view of the world“ (Goodchild 1995, S. 37) ist damit die technische Lösung des praktischen Problems, unterschiedliche Daten zu verwalten, die – bezogen auf die konkreten Fragestellungen in inhaltlich sinnvoller Weise – Punkten, Linien oder Flächen, d.h. Teilmengen der Erdoberfläche zugeordnet werden sollen. Die Repräsentation einer diskreten Menge sogenannter ‘räumlicher Objekte’ in Form von Layern erfolgt dabei mit Hilfe unterschiedlicher Datenmodelle: als Punktmengen (z.B. Wetterstationen), reguläre Gitter (z.B. digitale Bilder in der Fernerkundung), homogene Areale (z.B. Bodenkarten), Linien gleicher Attributwerte (z.B. Höhenlinien), Graphen (z.B. Wegenetze) oder als Mengen irregulärer Dreiecke (als *triangulated irregular network* bei der Modellierung von Oberflächen). Entscheidend ist, daß alle Objekte eines Layers nebeneinander angeordnet sind, d.h. innerhalb *eines* Layers existieren keine zwei Objekte, die auf derselben Erdstelle verortet sind; jeder Layer stellt damit eine planare Subdivision der Erdoberfläche dar (vgl. jedoch unten sowie die Einleitung von Kap. 3). Genau das wird durch die themenbezogene Gliederung erreicht, denn wo eine Gemeinde ist, kann keine zweite sein (wohl aber ein Bundesland), wo eine Zuckerrübe wächst, kann keine zweite wachsen (wohl aber Schwarzerde vorkommen), und wo es regnet, kann es nicht trocken sein (wohl aber kalt).

Anhand der unterschiedlichen Datenmodelle wird zunächst grob zwischen sogenannten *Raster-* und *Vektor-GIS* unterschieden. Das von einem Raster-GIS unterstützte Datenformat ist das eines regulären Rasters bzw. Gitters aus regelmäßigen, rechteckigen Rasterzellen. Alle Layer einer Datenbank haben eine identische räumliche Ausdehnung, Aufteilung und Orientierung, so daß Zellen eines Layers in ihrer Lage und Größe mit den Zellen der anderen Layer vollständig übereinstimmen. Raster-GIS werden überwiegend in der digitalen Fernerkundung eingesetzt, da sie das Datenformat digitaler Bilder, d.h. geordnete Mengen rechteckiger ‘Bildpunkte’ bzw. ‘Pixel’ perfekt unterstützen. Prominente Vertreter dieser GIS-Gattung sind z.B. IDRISI, ein PC-basiertes System der *Clark University*, oder GRASS, ein Unix-basiertes System des *U.S. Army Corps for Engineers*.

Jeder Layer eines Vektor-GIS ist dagegen aufzufassen als Menge von irregulären Punkten, (Kontur-)Linien oder Arealen. Punkte sind durch zwei (in dreidimensionalen Anwendungen natürlich drei) Koordinaten, Linien als gerade Verbindung zweier Punkte und Flächen durch die geordnete Menge der sie begrenzenden Linien repräsentiert. Die geometrischen Modelle z.B. von gewundenen Flüssen oder Straßen werden in der Regel als Linienzüge (auch ‘Kantenzüge’ oder ‘Polylinien’), die von Arealen als Polygone approximiert. Das wohl populärste Vektor-GIS ist ARC/INFO der Firma ESRI (*Environmental Systems Research Institute*). Vektor-GIS erlauben ein sehr hohes Maß an geometrischer Präzision und werden daher neben den wissenschaftlichen Anwendungen häufig in technischen Zusammenhängen eingesetzt, bei denen es besonders auf präzise erdräumliche Koordinaten ankommt (z.B. Instandhaltung und Planung von Kanalnetzen). Ein solcher, rein (bau)technisch motivierter Ein-

<sup>4</sup>Im englischen Sprachgebrauch wird häufig auch von *themes* oder *coverages* gesprochen; im mathematischen Sinne handelt es sich um *Felder*.

satz von GIS führte zu einer Klasse von Vektor-GIS, bei der das Layer-(Feld-)Modell nicht stringent durchgehalten wird. In solchen Systemen fungieren Layer als Kollektionen beliebiger Objekte, die sich auch innerhalb eines Layers überschneiden dürfen (z.B. System/9 von *Computervision*).

GIS wird heute also in zahlreichen und sehr unterschiedlichen Anwendungszusammenhängen eingesetzt:

- Im wissenschaftlichen Kontext sind überwiegend problemspezifische raumbezogene Analysemethoden, d.h. Methoden der sogenannten *Spatial Analysis* gefragt (vgl. v.a. Berry, Marble 1968). GIS stellt diese Methoden zur Verfügung, z.B. in Form von Regionalisierungsverfahren, Gravitations- und Lokations-Allokationsmodellen, Routensuch- und -optimierungsverfahren sowie weiteren raumbezogenen Funktionen, z.B. Verschneidungsverfahren (geometrische Kombinationen verschiedener Layer) oder Pufferungen (Bestimmung von Bereichen maximaler Entfernung) u.v.m.
- Der Einsatz von GIS in der administrativen oder räumlichen Planung bietet die Möglichkeit, schnell, einfach und gezielt Informationen für die jeweiligen am Planungsprozeß beteiligten Personen bzw. Institutionen zusammenzustellen. Die Anwendungsbereiche reichen von einfachen Informationsabfragen und kartographischen Darstellungen über die Bereitstellung von Diskussionsvorlagen oder die streng formalisierte Erstellung von Kriterienkatalogen bei der Flächennutzungsplanung (z.B. bei der Bestimmung der 'ökologischen Wertigkeit' von Flächen) bis hin zur zentimetergenauen Positionierung von Kanalrohren und Straßenbahnschienen oder der Analyse von Regionalstatistiken.
- Zu GIS im 'Alltag' zählen schließlich Hotelbuchungssysteme, Bürgerinformationssysteme, Verkehrsleitsysteme und ähnliche Systeme, die es GIS-unerfahrenen Anwendern ermöglichen, sich räumlich zu orientieren (vgl. z.B. Garner 1993, Ponn 1996). Für die meisten Anwendungsfälle im Alltag ist die Exaktheit des geometrischen Bezugssystems vollkommen unbedeutend, z.B. wenn ein städtisches Straßennetz auf dem Display des Satelliten-Navigationsgerätes eines Fahrzeugs auf wenige, grob die relative Fahrtrichtung symbolisierende Pfeile reduziert wird.

Während Arbeitsergebnisse im wissenschaftlichen Bereich häufig auf einem einheitlichen Sprach- und Wissensniveau diskutiert werden können, kommt im planerischen Kontext Aspekten des Informationsaustausches zwischen unterschiedlich vorinformierten Instanzen eine größere Bedeutung zu. Deshalb liegt der Schwerpunkt des mit diesen Informationssystemen implementierten Methodenspektrums weniger auf analytischen Techniken, sondern eher auf Routinen zur Aggregation und (karto-)graphischen Aufbereitung von Informationen (vgl. die Beiträge in Dollinger, Strobl 1989b bis 1997, Fischer, Nijkamp 1993b; Moll 1995; auch Couclelis 1991; Anselin, Getis 1993). Die Wahl des geometrischen Bezugssystems ist dabei stark vom jeweiligen Kontext abhängig: Die Sanierung eines Abwassernetzes erfordert beispielsweise detailgenaues, großmaßstäbiges und lagetreues Kartenmaterial, während bei der Erstellung und Präsentation großräumiger planerischer Leitbilder eher nicht unbedingt maßstabgetreue Kartenskizzen Verwendung finden, bei denen die Wiedergabe grober und geometrisch häufig extrem ungenauer räumlicher Strukturen im Vordergrund steht. Für Navigationssysteme stehen schließlich Informationsabfrage- und Visualisierungsfunktionen im Zentrum des Interesses.

## 1.2 GIS als Geographischer Informations-Szientismus

Es ergibt sich in etwa folgendes Bild: Je nach Anwendungszusammenhang werden die Anforderungen an GIS<sup>5</sup> unterschiedlich gewichtet, und ein GIS ist mal als wissenschaftlicher Werkzeugkasten, mal als organisationsorientierte Verfahrenshilfe und mal als schlichtes Informationsmedium bzw. als Orientierungshilfe aufzufassen. In dieser Folge nimmt die Bedeutung von Analysetechniken ab, die von Visualisierungstechniken dagegen zu. Entsprechend ausdifferenziert ist auch die Liste der angebotenen Systeme: Bodeninformationssysteme, Hochwasserinformationssysteme, Landinformationssysteme, Landschaftsinformationssysteme, Leitungsinformationssysteme, Regionalstatistische Informationssysteme, Umweltinformationssysteme, Verkehrsinformationssysteme etc. (vgl. ausführlich Dollinger, Strobl 1989a).

Die Tatsache, daß all diese Hard- und Software-Varianten trotz ihrer offensichtlichen Verschiedenheit unter dem gemeinsamen Begriff 'Geographisches Informationssystem' (für Nicht-Geographen häufig auch 'Geo-Informationssystem' oder 'Raumbezogenes Informationssystem') zusammengefaßt werden, berührt GIS-Anwender zunächst natürlich kaum. Auch als GIS-Entwickler kann man sich auf den Standpunkt stellen, immer genau das zu entwickeln und als GIS (besser: BIS, HWIS, LIS, RIS, UIS usw.) zu verkaufen, was sich Kunden und Endanwender eben gerade darunter vorstellen. Eine solche pragmatische Einstellung ist mit Sicherheit praxisnah, lukrativ und selbstverständlich legitim und macht die weitere theoretische Diskussion überflüssig.

Eine gänzlich andere Situation ergibt sich allerdings, wenn aus einer wissenschaftlichen – und insbesondere geographischen – Perspektive Fragen nach den theoretischen oder wissenschaftlichen Grundlagen von GIS gestellt werden. So lautet z.B. der Untertitel der im zweijährigen Rhythmus stattfindenden Conference on Spatial Information Theory programmatisch „A theoretical basis for GIS“ (z.B. Frank, Kuhn 1995). Maguire sieht in GIS neben einer geographisch ausgerichteten Computertechnologie „a new *discipline* which is generating massive interest world wide“ (Maguire 1991, S. 9; Hervorhebung von N.M.), die insbesondere auf raumbezogene Analyse und Modellbildung fokussiert ist, „in which GIS is seen more as a spatial information *science* than a technology“ (ebd., S. 14; Hervorhebung von N.M.). Analog schlägt Goodchild vor, das Akronym 'GIS' als „geographic information *science*“ zu dekodieren und verweist auf „agendas for research *on* GIS, not research *with* GIS, although clearly the first advances the second goal“ (Goodchild 1995, S. 42; Hervorhebungen im Original).

Die anvisierte 'Verwissenschaftlichung' von GIS als Forschungsobjekt erfolgt in mehreren Spielarten. In der inzwischen bereits zum Klassiker avancierten Aufsatzsammlung von Maguire, Goodchild und Rhind (1991a) werden fünf Schlüsselprobleme („key scientific questions“; dies. 1991c, S. 111) herausgestellt, denen sich eine akademische GIS-Disziplin annehmen soll (vgl. auch Laurini, Thompson 1992, S. 20/21):

- *nature of spatial data*, also die besondere Charakteristik erdraumbezogener im Gegensatz zu nicht-erdraumbezogenen Daten,
- *digital representation*, d.h. Fragen nach technischen Aspekten der Computer Hard- und Software wie z.B. Datenmodelle, Datenbanktechniken, effiziente Datenabfragestrategien und andere informationstechnische Details,
- *functional issues*, z.B. die Klärung der sogenannten geographischen Primitiven (geographisch bzw. geometrischen Grundeinheiten) oder der Umgang mit Fehlern und Unsicherheiten bei raumbezogenen Daten und deren Effekte auf eine raumbezogene Analyse und Modellbildung,

<sup>5</sup>Ungeachtet der einzelnen Kapitelüberschriften wird das Akronym 'GIS' im Text immer als 'Geographisches Informationssystem' dekodiert (vgl. auch Fn. 1).

- *display issues*, mit denen die Aufdeckung der Beziehungen zwischen Karten, Tabellen oder Diagrammen und den menschlichen Wahrnehmungs- und Interpretationsleistungen gemeint sind und
- *operational issues*, die Aspekte der GIS-Aquisition und Installation bei Endanwendern und dabei entstehende juristische Fragen, aber auch die Entwicklung von Datenbankstandards umfassen (ebd., S. 111ff.).

In der Tat ist es dem Herausgeberteam mit dieser Sammlung von Schlüsselproblemen gelungen, die konkreten Problemfelder zu umreißen, mit denen sich die GIS-Gemeinde konkret beschäftigt. Im folgenden soll untersucht werden, inwieweit sich daraus *Forschungsfragen* für die GIS-Gemeinde ergeben und wie es um deren *wissenschaftlichen* Gehalt bestellt ist.

### 1.2.1 Die Charakteristik erdraumbezogener Daten

Die besondere Charakteristik erdraumbezogener Daten wird in der modernen Geographie vorrangig unter dem Begriff der *räumlichen Autokorrelation* diskutiert: Zwei Ereignisse an den Orten (Erdstellen) *A* und *B* korrelieren in der Regel um so stärker miteinander, je kleiner die Distanz zwischen *A* und *B* ist. Oder wie Tobler (1970) es als sein „first law of geography“ ausdrückt: „Everything is related to everything else but near things are more related than distant things“. Analog formuliert Bartels das Koinzidenzprinzip, nach dem das Auftreten verschiedener Phänomene an derselben Erdstelle eine Ursache-Wirkung-Interpretation geradezu provoziert, denn „wo . . . räumliche Koinzidenz gegeben ist, liegen stets Hypothesen eines inhaltlichen Zusammenhangs nahe“ (Bartels 1970a, S. 16). Tobler und Bartels bringen damit einen der (wenn nicht den) Grundpfeiler eines jeden geographischen Erkenntnisgewinnungsprozesses auf den Punkt. Entsprechend wirbt das Softwarehaus ESRI als einer der Marktführer für seine Spatial Database Engine mit dem Slogan „Once You See *Where* . . . You’ll Understand *Why*“ (z.B. in GIS-Europe 1997/12, S. 51; Auslassung und Hervorhebungen im Original).

Eine weitere Charakteristik liegt im Auftreten sogenannter *ökologischer Fehler* (vgl. Robinson 1950; Openshaw 1978; Johnston 1980, S. 263f.; Bahrenberg, Giese, Nipper 1985, S. 198f.). Während bei Vorhandensein räumlicher Autokorrelation zwei beliebige Variable (in der Regel) um so stärker korrelieren, je näher die Erdstellen zueinander liegen, an denen sie erhoben wurden, liegen ökologische Korrelationen in der Größe und Form der räumlichen Einheiten begründet, in denen bzw. für die die Variablen erhoben wurden. Insbesondere treten ökologische Fehlschlüsse auf, wenn aufgrund von Korrelationen innerhalb eines räumlichen Bezugssystems – z.B. die Beobachtung, daß hohe (durchschnittliche) Kriminalitätsraten in Arealen mit hoher (durchschnittlicher) Armut auftreten – Rückschlüsse auf die Korrelation derselben Variablen auf einem höher auflösenden räumlichen Bezugssystem oder gar auf der Individualebene gezogen werden – z.B. der Schluß, daß arme Personen mehr Verbrechen begehen als reiche. Ein ähnlicher Effekt tritt innerhalb desselben Bezugssystems auf, wenn die Form der räumlichen Basiseinheiten variiert wird (das sogenannte *modifiable area unit problem*; vgl. Gehlke, Biehl 1934; Neprash 1934; Fotheringham, Wong 1991; Fotheringham, Densham, Curtis 1995).

### 1.2.2 Spatial Analysis

Aufgrund dieser Besonderheiten erdraumbezogener Daten fordern zahlreiche Autoren (unter ihnen vor allem die quantitativen Geographen) die Integration von GIS und den Methoden der *Spatial Analysis* (z.B. Berry, Marble 1968; Martin 1991; Openshaw 1991; Ding, Fotheringham 1992; Anselin, Getis 1993; Fischer, Nijkamp 1993b; Fotheringham, Rogerson 1993, 1994a, 1994b; Harris, Batty 1993; Nijkamp, Scholten 1993; Openshaw, Scholten 1993; Densham 1994; Fischer, Scholten, Unwin 1996; Unwin 1996). Unter diesem Begriff werden alle Verfahren subsumiert, die Gattrell als „application of

statistical methods to the solution of geographical research questions“ beschreibt (Gartrell 1987; zit.n. Anselin, Getis 1993, S. 36). Gemeint sind Regionalisierungsverfahren, Lokations-Allokationsmodelle, demographische Modelle, raumbezogene Autokorrelationsmaße, Gravitationsmodelle u.v.m., d.h. Verfahren, die größtenteils im Zuge der sogenannten *quantitativen Revolution* entwickelt bzw. von der Geographie adaptiert wurden und die sowohl mit räumlicher Autokorrelation als auch mit ökologischen Fehlern umzugehen vermögen.<sup>6</sup>

Seit einiger Zeit wird auch die Implementierung explorativer Analysemethoden im Sinne einer Spatial Analysis diskutiert (z.B. Fotheringham, Rogerson 1992; Openshaw, Scholten 1993; Openshaw 1994; Fischer, Scholten, Unwin 1996). Explorative Methoden sollten v.a. in den Bereichen zur Anwendung kommen, in denen formalmethodisch wenig oder nicht ausgebildete Entscheidungsträger mit GIS arbeiten und GIS als Instrument zur Entscheidungsunterstützung einsetzen. Der Begriff der ‘explorativen Datenanalyse’ bezeichnet dabei alle Arten der Datenvisualisierung (Tukey 1977), aber auch automatisierte, induktive Methoden der Mustererkennung, wie z.B. clusteranalytische Verfahren (Openshaw 1990). In einem etwas engeren Sinne und als Abgrenzung zu numerischen Methoden wird der Begriff der ‘explorativen Datenanalyse’ häufig allein im Sinne von ‘Datenvisualisierung’ gebraucht. In jedem Fall wird einer explorativen Analyse sowohl eine vorbereitende Funktion für weiterführende (quantitative) Analysen als auch eine eigenständige Analysefunktion zugesprochen.

Das Grundproblem dieser Art raumanalytisch-zentrierter Geographie ist im wissenschaftstheoretischen Diskurs des Fachs nur zu gut bekannt: Die Tatsache, daß z.B. aufgrund eines geographischen (distanzabhängigen) Modells eine richtige Vorhersage getroffen wird, bedeutet nicht, daß das vorhergesagte Ereignis ursächlich auf die Distanzvariable zurückzuführen ist; „Empiricist-positivist geography and technical control equate prediction with explanation“ (Taylor, Johnston 1995, 55; vgl auch Johnston 1994). Mit ihrer ausdrücklichen Betonung der Modellbildung und Analyse, die neben der reinen Betrachtung also insbesondere auf das Verständnis, die Interpretation und die Erklärung (v.a. auch sozialer Phänomene und Prozesse) abzielt, läuft der quantitativ orientierte Teil der GIS-Gemeinde also Gefahr, in eben jenen „diffusen Geodeterminismus“ abzurutschen, den Hard bereits vor fast zwei Jahrzehnten als Absurdität entlarvt hat: „Die ‘naturgeographischen’ Faktoren und der Faktor ‘Distanz’ bzw. ‘Lage’ erklären nicht nur nicht alles . . ., sie erklären für sich allein *gar nichts*“. Erdräumbezogene Größen, so argumentiert Hard, „sind nicht Erklärungsgrößen *neben* anderen (z.B. sozialwissenschaftlichen) Erklärungsgrößen; denn sie erklären nur etwas im Zusammenhang mit andern, ‘nicht-geographischen’ Variablen *und* dann, wenn sie in ‘nicht-geographische’ Variablen übersetzt werden“ (Hard 1982b, S. 106; Hervorhebungen im Original).

Mit anderen Worten: Man kann (fast) alles mit erdräumlichen Koordinaten versehen, unabhängig davon, ob dies sinnvoll ist oder nicht – eine erdräumliche Projektion führt allerdings nur selten zu einem relevanten Erkenntnisgewinn. So mag beispielsweise die Email-Korrespondenz des Autors der vorliegenden Schrift Freunde sowohl aus seiner unmittelbaren Bremer Nachbarschaft als auch aus Schweden oder Australien umfassen. Eine erdräumliche Kartierung dieses persönlichen ‘Kommunikationsraumes’ wäre als Beschreibung bestenfalls belanglos; als Erklärung des Kommunikationsverhaltens wäre sie als Ausdruck der „blickverengende[n] Unfruchtbarkeit einer chorologischen Brille“ (Bahrenberg 1987, S. 234) geradezu naiv. Sinnvoller wäre vielmehr, den Kommunikationsraum z.B. im Sinne der technischen Erreichbarkeit zu beschreiben. So wird die Kommunikationsform und -häufigkeit beispielsweise davon beeinflusst, ob der Autor den nächsten Netzknoten nur über ein Modem

---

<sup>6</sup>Mit der quantitativen Revolution wurde seit Beginn der 60er Jahre der Versuch unternommen, Geographie als Wissenschaft neu zu etablieren. An die Stelle der traditionellen Länderkunde mit ihrer deskriptiv-qualitativen Methodik sollte eine moderne Wissenschaft treten, die auf einer mathematisch fundierten Theorie- und Modellbildung basierte und auf ‘gesellschaftlich relevante’ Fragen ausgerichtet war. In dieser Zeit wurde der Großteil des geographisch-quantitativen Instrumentariums entwickelt, dessen Hauptwesenszug die Einbindung räumlicher Größen (z.B. Distanz) als ‘erklärende’ Variable in der Theorie- und Modellbildung ist.

mit langwieriger Einwahlprozedur und kostspieliger 'Online-Zeit' erreichen kann, oder ob er direkt an ein Hochgeschwindigkeitsnetz angeschlossen ist. Eine solche Kodierung wäre keine erdräumliche mehr, obwohl durchaus geometrisch in Form von Bildern, Diagrammen u.ä. beschreibbar. Mit der Übersetzung 'geographischer' in 'nicht-geographische' Variablen (z.B. 'Natur' durch 'wahrgenommene Natur', 'Distanz' durch 'Transportkosten' oder 'Kommunikationswiderstände') wird natürlich jedes geographische Problem in ein nicht-'raum'wissenschaftliches Problem transformiert, welches dann nicht mehr sinnvoll mit GIS bearbeitet werden kann (außer es wird erneut 'verräumlicht').

Ungeachtet der wissenschaftstheoretischen Bedenken ist die Freude der quantitativen Geographen groß, die mit GIS nun endlich die Möglichkeit zu haben scheinen, ihre rechenintensiven Analyse- und Modellierungsverfahren auf die geographische Welt anzuwenden: „The opportunity in the use of GIS is that it indeed has made previously prohibitive computationally intensive and highly visual ways of spatial analysis accessible at reasonable costs“ (Anselin, Getis 1993, S. 46). Insbesondere der technische Fortschritt bei der Computerhardware läßt hoffen, daß Spatial Analysis bald überall und von jedem durchgeführt werden kann, denn „we are rapidly approaching the time when every desktop PC will be able to perform spatial analysis“ (Unwin 1996, S. 540). Die Forschungsinhalte einer quantitativen GIS-Geographie bestehen daher im wesentlichen darin, die Methoden der Spatial Analysis weiter zu verfeinern und in GIS zu implementieren. Ersteres ist Aufgabe der quantitativen Geographie, für letzteres bedient man sich der jungen Disziplin der Geoinformatik.

### 1.2.3 Geoinformatik

Für Geoinformatiker ist das natürlich alles andere als ein vielversprechendes Forschungsprogramm, und so herrscht eher Skepsis: „Also doch nichts neues? Ist GIS-Technologie nichts weiter als der Versuch, lang erprobte Methoden der Visualisierung und Interpretation EDV-tauglich zu machen?“ (Bartelme 1995, S. 30). Keineswegs, denn Geoinformatik, die ja immerhin „Einzug in die Fakultäten, die Fachsymposien, die Literatur“ (ders. S. 28) hält, muß mehr sein als alter Wein in neuen Schläuchen. Geoinformatik wird daher gerne zum Ausdruck einer Akademisierung der GIS-Gemeinde stilisiert: „Heute macht man bereits den nächsten Schritt von der Technologie zur Wissenschaft. . . . Man erkennt den Nutzen, den eine Symbiose zwischen der anwendungsorientierten Technologie und einer tiefer schürfenden wissenschaftlichen Auseinandersetzung mit den Grundlagen, also dem Raumbezug an sich, mit der Parametrisierbarkeit und mit Analysemodellen bringt“ (ebd.).

Dabei soll es nicht nur um die eigentliche informationstechnische Aufgabe gehen, das analoge Informationsmedium 'Karte' oder 'Plan' in ein digitales Medium zu transformieren. Im Zentrum des geoinformationswissenschaftlichen Interesses steht vielmehr die Fähigkeit der Betrachter und Nutzer dieser Medien, „durch assoziatives Denken raumbezogene, strukturelle und darstellungsspezifische Merkmale zu interpretieren und miteinander in Beziehung zu setzen“ (ders. S. 30). Offensichtlich liegt hier der erwartete symbiotische Nutzen, und der „Raumbezug an sich“ als Dreh- und Angelpunkt einer jeden „tiefer schürfenden wissenschaftlichen Auseinandersetzung“ fördert schließlich folgendes Forschungsprogramm zu Tage: „Derartige Abläufe im menschlichen Denkprozeß besser verstehen zu lernen und Modelle für die Entscheidungsfindung in Situationen mit raumbezogenem Charakter zu finden, das Speichern, Wiedererkennen und Weitervermitteln solcher Situationen in gewissem Umfang nachvollziehen zu können, das ist die eigentliche Herausforderung an die Geoinformatik, die damit den Rang einer *vollwertigen und eigenständigen Wissenschaftsdisziplin* erhält“ (ebd.; Hervorhebung von N.M.).

Mit dem, was hier als Forschungsprogramm einer eigenständigen Geoinformatik vorgeschlagen wird, nämlich ein Verständnis für Speicher-, Wiedererkennung- und Weitervermittlungstechniken 'raumbezogener Situationen' zu entwickeln (zumindest in gewissem Umfang!), bleibt zunächst offen,

was „Situationen mit [erd]raumbezogenem Charakter“ eigentlich sind (im Gegensatz zu Situationen *ohne* erdraumbezogenen Charakter). Für eine Geoinformatik, die sich auf den „[Erd]Raumbezug an sich“ beruft, sind solche Differenzierungen anscheinend trivial. Blickt man auf die Inhalte, die die Geoinformatik vermitteln will, wird zudem klar, daß es in Wirklichkeit um ganz etwas anderes geht als um menschliche Denkprozesse, Wissensvermittlung u.ä., nämlich um Datenmodelle, Datenorganisation, Datenbanktechnik, graphische Benutzeroberflächen, Applikationen u.v.m., eben um jene technischen Bereiche, die im wesentlichen seitens der Informatik behandelt werden (vgl. auch Schilcher 1991).

Darüber hinaus, und für das Fach offensichtlich von wesentlicherer Bedeutung als die Forschungsinhalte, sieht sich die neue Disziplin als Konglomerat verschiedener Fachrichtungen. Es gibt „(noch) keine ‘Original-Informatiker’, sondern Geodäten, Geographen, Kartographen, Geowissenschaftler, die ihre Disziplin entsprechend erweitert haben und im Kontext der neuen Technologie sehen“ (Bartelme 1995, S. 28). Es wird gar nicht erst versucht, die GIS-Technologie aus einem disziplinären Kontext heraus zu sehen, vielmehr besteht die multisziplinäre Vollwertigkeit der Geoinformatik als Wissenschaftsdisziplin darin, einzelne Disziplinen im gemeinsamen Kontext einer alles verbindenden Technologie zu begreifen! Dies führt beinahe zwangsläufig dazu, daß die propagierte ‘tiefer schürfende wissenschaftliche[] Auseinandersetzung’ jenen einführenden und oberflächlichen Charakter annimmt, der typisch ist für die geoinformatische Literatur. Beispielhaft mag der Anspruch von Laurini, Thompson stehen, den sie für ihr Lehrbuch deklamieren: „... we have aimed to reach both academician and practitioner, both system designer and database manager, and the intuitive and formal learner“ (Laurini, Thompson 1995, S. xvi). Eine Geoinformatik, die allen alles erklären will, schwappt zwangsläufig in einer oberflächlich-wissenschaftlichen Grauzone, und der Anspruch auf Eigenständigkeit der Wissenschaftsdisziplin führt weniger zu einer vollwertigen Geo-, sondern eher zu einer (erd)oberflächlichen Vollwert-Informatik.

#### 1.2.4 Spatial Information Theory

Ein weiterer Teil der akademisierten GIS-Gemeinde, der sich unter dem Label der ‘Spatial Information Theory’ versammelt, „brings together three fields of research of enormous importance for the Geographic Information System (GIS) technology: spatial reasoning, representation of space, and human understanding of space“ (Frank, Kuhn 1995, S. v). Mit *Spatial Reasoning* sind menschliche Deduktionsleistungen gemeint, d.h. „the logical deduction of spatial information from spatial facts“ (ebd.). *Representation of Space* steht in der Tradition der Künstlichen Intelligenz-Forschung, „but is also practically important for the design of spatial data structures“ (ebd.) und *Human Spatial Cognition* fokussiert auf die menschlichen Fähigkeiten, reale (‘räumliche’) Objekte (das sind die „spatial facts“) wahrzunehmen, zu unterscheiden und sich an ihnen auf der Erdoberfläche zu orientieren, auf raumbezogene Aspekte der Sprache und „the cooperating fields of philosophy (spatial ontology), linguistics, and psychology“ (ebd.).

Auch bei dieser GIS-wissenschaftlichen Variante geht es also im wesentlichen um die Anwendung sowohl geographischer als auch nicht-geographischer Erkenntnisse für das Design von GIS. Einen der Grundgedanken drücken z.B. Egenhofer und Mark mit ihrer „Naive Geography“ aus, „the body of knowledge that people have about the surrounding geographic world“, die für sie den Ausgangspunkt einer Theoriebildung darstellt „for designing future Geographic Information Systems that follow human intuition and are, therefore, easily accessible to a large range of users“ (Egenhofer, Mark 1995, S. 1). Daneben werden Metaphern (Kuhn 1993), kognitive Bildschemata (Mark 1989; Kuhn, Frank 1991), formalisierte Konzepte des ‘geographischen Raums’ (Mark, Frank 1990; Bittner, Frank 1997), Navigationsmechanismen (Dieberger 1995; Hanhwe, Hirtle 1995), ‘raumbezogene’ Kognition (Medyckyj-Scott, Blades 1992), ‘raumbezogener’ Wissenserwerb (Montello, Freudenschuh 1995)



oder natürliche Sprachen (Frank, Mark 1991) auf ihre Anwendbarkeit für das Design von GIS (in Form von Benutzeroberflächen, Datenbanken, Abfragesprachen etc.) geprüft (vgl. auch Nyerges et al. 1995).

Eine Spatial Information *Theory* ist da nur schwer zu entdecken. Es geht eben doch bloß um die praktische Umsetzung kognitionswissenschaftlicher (philosophischer, linguistischer, mathematischer, psychologischer, geographischer u.v.m.) Erkenntnisse auf GIS-Design und -Technik und nicht – wie das etwas hochgestochene Label zunächst vermuten läßt – um eine einheitliche raumbezogene Theorie (was sollte eine solche Theorie auch thematisieren?). Der theoretische Gehalt dieser Form von akademischer GIS-Arbeit wird vielmehr aus zahlreichen (beinahe beliebigen, immer aber ‘raumwissenschaftlichen’) Disziplinen assimiliert. Die ‘theoretische Basis für GIS’ ist demnach nichts weiter als die Summe aller (‘raum’)wissenschaftlichen Theorien, die eingesetzt werden für das Design eines (geographischen) Werkzeugs, dessen Anwendungsrelevanz außer Frage steht. Mit anderen Worten: Es geht nicht um das Forschen *über* GIS, sondern um das Forschen *für* GIS, und das führt zwangsläufig zu einem Sammelbecken wissenschaftlicher und technischer Disziplinen, in dem sich Informatiker, Linguisten, Kartographen, Logiker, Kognitionspsychologen, Fernerkundler, Philosophen, Geodäten, Verhaltenspsychologen, Statistiker, Ethnologen und neben weiteren schließlich auch Geographen tummeln.

### 1.3 GIS als Geographisches Integrations-System

Der wissenschaftliche Gehalt der skizzierten Problemfelder im Sinne einer ‘research on GIS’, d.h. im Sinne eines Forschungsprogramms, das GIS selbst zum Objekt der Forschung macht, fällt ganz offensichtlich recht mager aus. Von einer solchen Forschung sollte man erwarten, daß sie Fragen nach dem Warum stellt, also beispielsweise versucht, die (nicht-technische) Funktionsweise oder den Erfolg von GIS (besonders auch außerhalb der Geographie) zu *erklären*. Stattdessen werden quantitative Methoden in neue technische Kleider gewandet, werden neue akademische Disziplinen oder raumtheoretische Vereinigungsmengen heraufbeschworen und unter wissenschaftlichen Labels verkauft. Das Streben der GIS-Gemeinde nach Akademisierung, Abgrenzung und Eigenständigkeit kann jedoch nicht darüber hinwegtäuschen, daß es immer um die technische Perfektionierung des Werkzeuges GIS geht und immer um technische statt um inhaltliche Erklärungen. Die GIS-Gemeinde betreibt damit eher solides Handwerk als (theoretische) Wissenschaft.

Hier und da ist auch ein gesunder Pragmatismus zu erkennen, z.B. wenn Worboys bemerkt: „One way of understanding GIS is to look at its application“ (Worboys 1995, S.2). In dieser Einsicht liegen erstaunliche Parallelen zum geographischen Selbstverständnis, war es doch „... während der 2000-jährigen Geschichte der Geographie wohl meistens so, daß die Arbeitspraxis der Theorie vorausgegangen ist ...“ (Bobeck 1957, S. 122). Die akademische GIS-Gemeinde führt dieses geographietheoretische Lamento beinahe nahtlos fort, denn „... the term GIS has continued to be defined more by its applications and technology than by any clear principles, or any strong sense of an intellectual or conceptual core“ (Maguire, Goodchild, Rhind 1991c, S. 111).

Die Identifikation der GIS-Gemeinde über ihr gemeinsames Tun steht damit in jahrtausendealter geographischer Tradition, ist aber auch konform zu jüngeren fachhistorischen Entwicklungen. Mit dem Aufkeimen von Fragen nach der (sozialen) Relevanz der Theorien und der Anwendbarkeit quantitativer Methoden zur Lösung gesellschaftspolitischer Mißstände verlagerte sich nämlich der fachliche Schwerpunkt in den 70er Jahren von der abstrakten Modellbildung der ‘quantitativen’ hin zu einer ‘angewandten’ Geographie. Der Brückenschlag zwischen abstrakter räumlicher Theoriebildung und anwendbaren Modellen und Methoden wurde v.a. durch die Hinwendung zur systemanalytischen Forschung und Modellbildung vollzogen. In der Betonung der Anwendbarkeit geographischer Forschung wird heute (zumindest seitens der ‘politischen Geographie’) einer der Hauptgründe für den Erfolg von

GIS innerhalb des Fachs gesehen: „The proponents of GIS combine the early technical concerns of quantifiers with the later social, economic, and political concerns of those who advocate applied geography. Hence, although they derive their positions from trends set in motion by the quantitative revolution, as a major tendency in contemporary geography GIS practitioners constitute a new grouping of ideas“ (Taylor, Johnston 1995, S. 54). Mit der Betonung des anwendungsrelevanten analytischen Potentials von GIS verschwimmen also schließlich die Grenzen zwischen Methodik (‘quantitativer Geographie’) und Anwendung (‘angewandter Geographie’). Damit stellt die GIS-Gemeinde den (vorläufigen) Höhepunkt der Entwicklung einer geographischen Spezies dar, die Taylor und Johnston als „strictly applied quantitative geographers“ bezeichnen (ebd.).

GIS ist das, was mit GIS gemacht wird, und deshalb zählt allein die Technik: „... . geography is infinitely complex and must be generalized, approximated or abstracted in order to be represented within the finite dimensions of a discrete computing device“ (Maguire, Goodchild, Rhind 1991c, S. 111/112). Die ganze Welt in ihrer unendlichen Komplexität wird digitalisiert, und die GIS-Gemeinde fragt nicht nach dem *Warum*, sondern nur nach dem *Wie*. Denn das *Warum* scheint außer Frage zu stehen, schließlich geht es um nichts weniger als um die Rettung der Menschheit: „The key forces which drive GIS include global, national and local questions of human survival in this earthly environment“ (Aangeenbrug 1991, S. 102).

Was aber mit GIS gemacht wird, ist eben auch genau das und nur das, was mit GIS gemacht wird, und das ist keineswegs beliebig! Niemand würde beispielsweise die Arbeit eines Radiologen, der mit Hilfe eines Computertomographen erkranktes Gewebe eines Organs zu erkennen sucht, als geographisch bzw. die eingesetzte Software als GIS bezeichnen, obwohl primär thematische Attribute (die Strahlentransparenz erkrankten Gewebes) in einem Raum (z.B. einem zweidimensionalen Schnitt des erkrankten Organs) mittels eines geometrischen Bezugssystems (einem kartesischen Koordinatensystem) verortet und auf (räumlich mit hoher Wahrscheinlichkeit autokorrelierte) Muster hin untersucht werden. Dies fällt (rein technisch) eindeutig unter den Begriff der Spatial Analysis, steht (rein technisch) nicht im Widerspruch zu gängigen GIS-Definitionen und dient auch der (zumindest partiellen) Rettung der Menschheit.

Eigenartigerweise bietet die GIS-Technik also auf der einen Seite das integrative Element der GIS-Gemeinde; auf der anderen Seite läßt sich GIS aber nicht über rein technische Kriterien von anderer, nicht als GIS bezeichneter Software unterscheiden. Die Integrationsleistung der GIS-Technik für die ‘Raum’wissenschaften liegt offensichtlich eher in ihrer Funktion als Vehikel des ‘Erdräumlichen’, d.h. die Technik als sinnstiftendes Element der akademischen GIS-Gemeinde ist wiederum selbst über die Erdoberfläche integriert. Zahlreiche Autoren benennen die Erdoberfläche (in der Regel in Vorworten und Festreden) ausdrücklich als einzig ‘gültige’ Projektionsfläche geographischer Daten. So sind für Tomlin die mit GIS zu bearbeitenden „facts“ ausschließlich erdoberflächlich: „A geographic information system is a facility for preparing, presenting, and interpreting facts that pertain to the surface of the earth“ (Tomlin 1990, S. xi). Diese Vorstellung finden wir auch in zeitgemäßer, d.h. umweltorientierter Fassung, z.B. bei Goodchild, Parks und Steyaert: „GIS by definition is a technology designed to capture, store, manipulate, analyze, and visualize the diverse sets of georeferences data that are required to support accurate modelling of the Earth’s environmental processes“ (1993, S. v). Die Welt ist für die GIS-Gemeinde gewissermaßen in ‘Umwelten’ parzelliert, die räumlich – als sich auf der Erdoberfläche verteilende Fakten – klar umrissen sowie in Form von georeferenzierten Datensätzen operationalisiert werden können und die mit GIS erhoben, gespeichert, manipuliert, analysiert, präsentiert, visualisiert und interpretiert werden sollen.

Die GIS-Gemeinde steht damit vor ähnlichen raum-begrifflichen Problemen wie die Geographie: „The terms ‘spatial’ and ‘geographical’ are often used interchangeably to describe geographical features. Strictly, the term spatial refers to any type of information about location . . . On the other hand,

geographical refers only to locational information about the surface or near surface of the earth at real-world scales and in real world space“ (Maguire 1991, S. 12). Mit GIS werden also nur Abbildungen und Phänomene betrachtet, die auf der Erdoberfläche oder Teilen davon existieren bzw. zu verorten sind und auch nur solche, deren räumliche Ausdehnung in einer Größenordnung zwischen wenigen Metern und dem Erdumfang liegen und damit dem „in der geographischen Forschung interessierenden Größenspektrum“ (Hagget 1973, S. 7)<sup>7</sup> entsprechen. In diesem Sinne ist auch der Widerspruch zu verstehen, der sich zunächst aufdrängt, wenn Laurini und Thompson in ihrem Standardlehrbuch über raumbezogene Informationssysteme schreiben: „Spatial is a term used here to refer to located data, for objects positioned in *any space, not just geographical*, a term we will use for the *world space*“ (Laurini, Thompson 1995, 4; Hervorhebung von N.M.). Auf den folgenden gut 600 Seiten wird zunächst klar, woraus dieser ‘world space’ besteht: aus Staaten, Städten, Flüssen, Tälern usw., eben Elementen ihrer sogenannten „real world“ (ebd., z.B. S. 5), die ganz offensichtlich doch durch und durch geographisch im o.g. Sinne ist. Der Verweis auf die Beliebigkeit des räumlichen Bezugssystems ist demnach in keiner Weise inhaltlich motiviert; die geometrische Verallgemeinerung ist jedoch notwendige Voraussetzung für eine abstrakte Diskussion der ‘geographischen’ Methoden, bei denen es sich wiederum ganz überwiegend um Methoden aus dem Bereich der Algorithmischen Geometrie handelt (vgl. O’Rourke 1994; Goodman, O’Rourke 1997).

Genauer über die Inhalte der mit GIS zu bearbeitenden Daten erfahren wir auch von Bartelme, für den ein Geoinformationssystem „der Erfassung, Speicherung, Analyse und Darstellung *aller* Daten, die einen Teil der Erdoberfläche und die darauf befindlichen technischen und administrativen Einrichtungen sowie geowissenschaftliche, ökonomische und ökologische Gegebenheiten beschreiben“, dient (Bartelme 1995, S. 12; Hervorhebung von N.M.). Was hier zunächst durchblickt, ist das traditionelle Kernthema der Geographie, nämlich „Konflikt und Harmonie von Mensch und Raum“ (Hard 1982c, S. 148). Die „facts that pertain to the surface of the earth“ sind nämlich nicht nur unbelebter und nicht einmal ausschließlich physischer Natur, sondern beinhalten auch „administrative Einrichtungen“ und „ökonomische Gegebenheiten“, d.h. soziale Einrichtungen und Prozesse, die – mit erdräumlichen Koordinaten versehen – Teil der geographischen ‘Umwelt’ und daher ebenso würdig sind, mit GIS bearbeitet zu werden. Und mehr noch: Auch scheinbar längst abgelegte geographische Methodiken erfahren mit GIS eine Renaissance, indem sich z.B. „das Schichtenkonzept . . . des länderkundlichen Schemas von Hettner (1932) sowohl als inhaltliche Inventarisierungsgrundlage als auch als Speicherkonzept (Layer-Struktur . . .) bewährt, der Flächenverschnitt und verschiedenste Formen der Überlagerung als die zentrale GIS-Methodik herausgebildet . . .“ haben (Margraf 1994b, S. 163).

Interessanterweise leben also mit GIS klassische geographische Sichtweisen wieder auf. Durch die Fokussierung auf die Erdoberfläche ist nämlich das Forschungsobjekt gewissermaßen vorgegeben: Fakten, Umwelten, Prozesse, die auf der Erdoberfläche oder auf Teilen von ihr aus geographischer Perspektive sinnvoll verortbar sind, werden auf Autokorrelation untersucht und nach dem Koinzidenzprinzip interpretiert. Auch wenn heute die Betonung auf Begriffen wie ‚Umwelt‘ oder ‚Prozeßorientierung‘ liegt, steht das Objekt, das mit GIS behandelt wird, demnach dennoch ganz in der Tradition der klassischen Geographie: „Es ist die Erdoberfläche im Sinne derjenigen Sphäre, in der sich das feste Land, das Wasser und die Luft berühren und teilweise durchdringen, einschließlich des Lebendigen, das sich in dieser Sphäre aufhält, darunter vor allem auch des Menschen, alles zusammen makroskopisch gesehen als irgendwie zusammengehöriges, ungemein mannigfaltig strukturiertes und in seiner räumlichen Erstreckung sowie im zeitlichen Ablauf inhaltlich abgewandeltes Ganzes. Dieses Objekt

<sup>7</sup>„Das in der geographischen Forschung interessierende Größenspektrum kann abgelesen werden an einem linearen Diagramm . . . Es besteht aus einer logarithmischen Skala, die von  $10^{30}$  bis  $10^{-50}$  cm reicht und damit die größten und die kleinsten je gemessenen Distanzen enthält (genauer  $10^{27}$  bzw.  $10^{-13}$  cm). In diesem Spektrum nehmen die Flächengrößen, mit welchen sich der Geograph beschäftigt, eine mittlere Position ein. Sie reichen von  $4,01 \cdot 10^9$  cm (Erdumfang) bis zu etwa  $10^3$  cm“ (Hagget 1973, S. 7).

kann im Ganzen oder auch in Teilen (räumlichen Ausschnitten) von der Geographie behandelt werden“ (Bobek 1957, S. 122). Eine moderne Interpretation dieser Auffassung gibt Göpfert, für den GIS ein Informationssystem ist „... das *alle* raumbezogenen Daten der Atmosphäre, der Erdoberfläche und der Lithosphäre enthält ...“ (Göpfert 1991, zit.n. Bartelme 1995, S. 12; Hervorhebung von N.M.).

Man beachte, daß das geographische Objekt damals wie heute immer als ein *Ganzes* behandelt werden soll, das höchstens in *Teilräume* untergliedert wird, um es einer geographischen Analyse leichter zugänglich zu machen: „Ob ‚Land‘ oder ‚Landschaft‘, stets handelt es sich aber um ein Stück Erdoberfläche in der ganzen Komplexheit seiner Erscheinungen“ (Bobek, Schmithüsen 1949, S. 114). Insbesondere gilt es, diese Ganzheit wissenschaftlich zu behandeln, und das bedeutet für Bobek „nicht nur Feststellung und Beschreibung der sinnlich wahrnehmbaren Gestalt des Objektes, sondern auch es zu erklären, d.h. die Kräfte und Bedingungen aufzudecken, unter deren Zusammenwirken es so geworden ist wie es sich darstellt“ (Bobek 1957, S. 122/123). Solange Geographen die Welt „makroskopisch sehen“, können sie immer das geographische Objekt, d.h. die „Ganzheit“ erblicken, in der alles „irgendwie zusammengehörig“ ist und die es zu behandeln gilt. Die geographische Interpretation dieser „Ganzheit“ ist die einer „sinnlich wahrnehmbaren Gestalt“, einer Gestalt, die sich gewissermaßen von selbst „darstellt“, d.h. ohne technische Hilfsmittel als konkretes Objekt rein sinnlich erfahrbar ist.<sup>8</sup> Es gilt, das „Wesen“ dieses Objektes aufzuklären, d.h. seine „Form und Struktur“, die „formenden (strukturbildenden) Prozesse“ sowie das „geschichtliche Werden“ (ebd. S. 123).

Offensichtlich wird genau dieses holistische Forschungsprogramm von der GIS-Gemeinde bewußt oder unbewußt adaptiert, allerdings in einer ‚modernen‘, d.h. auf die Informationstechnik zentrierten Form. In unseren modernen Zeiten zunehmender ‚Globalisierung‘ und drohender ‚Umweltkatastrophen‘ treibt die geographische Tradition wahre Stilblüten: „The complexity of the earth surface is such that in order to understand it a holistic view is essential. There is a need to use every available view and approach to study this complexity. GIS reflect our quest to combine all in order to see and predict better. GIS also allow, in fact demand, detailed analysis of a complex set of subjects of the environment. A clear understanding of the earth is still a long way off, but already scientists are better able to reflect on its complexity. It will require an enormous team effort of specialists, integrators, poets, politicians, and constant debate on how to do it better. The earth and its occupants demand it“ (Aangeenbrug 1991, S. 106).

Die auch hier wieder beschworene multisziplinäre GIS-Gemeinde läßt sich damit als Ausdruck einer klassischen geographischen Forschungslogik deuten, nach der nicht nur alles mit allem zusammenhängt, sondern nach der eine Reflexion der Komplexität dieses Allzusammenhangs am besten über die Projektion geeigneter Phänomen-Teilmenen auf die Erdoberfläche gelingt. Die mit GIS zu untersuchende Ganzheit ist damit Ausdruck der häufig beklagten Theorielosigkeit des Fachs (der Geographie als wissenschaftliche Disziplin). Analog zu einer Geographie, deren einziges sinnstiftendes Paradigma die ‚Räumlichkeit‘ ihres Untersuchungsgegenstandes ist, bleibt – wenn das einzige Kriterium zur Anwendung von GIS dasjenige ist, daß die zu speichernden, zu analysierenden, zu modellierenden, zu visualisierenden Daten mit Hilfe erdräumlicher Koordinaten kodierbar sind – schließlich kaum etwas übrig, was nicht mit GIS bearbeitet werden kann.<sup>9</sup>

Die gängige (geographische) Kategorisierung verfügbarer Software in GIS auf der einen und sonstiger Programme auf der anderen Seite ist damit keine notwendige Folge einer technischen Besonderheit von GIS. Nicht die Art, *wie* ein Problem behandelt werden soll (ob raum- bzw. geometriebezogen oder nicht), entscheidet über den Einsatz von GIS, sondern die Frage, *ob* ein Problem im geographi-

<sup>8</sup> „Die Geographie beschränkt sich ... nicht auf die Feststellung und Beschreibung des Sichtbaren, wengleich sie davon ausgeht“ (Bobek, Schmithüsen 1949, S. 112).

<sup>9</sup> Strenggenommen ist z.B. auch das vorliegende Schriftstück ‚räumlich‘: Breite und Höhe sind vom Deutschen Institut für Normung unter der Bezeichnung A4 mit 210 · 297 mm festgelegt, die dritte Dimension ist dagegen durch die gewählte Papierstärke sowie die Produktivität des Verfassers determiniert.

schen Sinne auf die Erdoberfläche projizierbar ist. Technisch ist der GIS-Begriff daher nicht kommunizierbar, und GIS-Entwickler stehen vor dem Widerspruch, daß das, was programmiert werden *soll* (ein Computersystem zur Erfassung, Verwaltung und Analyse großer Mengen *erdräumlich* verorteter Daten und darauf bezogener thematischer Attribute), niemals dem entsprechen wird, was programmiert werden *kann* (ein Computersystem zur Erfassung, Verwaltung und Analyse großer Mengen *geometrisch* verorteter Daten und darauf bezogener thematischer Attribute). Mit anderen Worten: Die geographische GIS-Idee ist faktisch, d.h. in Form einer Programmiersprache, nicht reproduzierbar.

Es ist jedoch ein leichtes, ungeachtet derartiger wissenschaftstheoretischer Absurditäten im verbindenden Interesse am und in den gemeinsamen Strukturen des ‚Räumlichen‘ eine der Existenzgrundlagen der Geographie zu sehen, indem vom erdoberflächlichen ‚Raum‘ über ‚räumliche Daten‘ erneut auf die GIS-Technik rekurriert wird: „. . . GIS itself rests on the assumption that different kinds of spatial data have common structures, and are processed in similar ways, and that there is consequently value in creating common spatial data handling and processing systems. And ultimately this is one argument for the existence of geography itself as a discipline“ (Goodchild 1995, S. 42).

Interessant ist also insbesondere die integrative Funktion des beforschten Objektes für das forschende Fach. So hatte die Geographie früher „mit den Landschaften als den natürlichen Regionen . . . endlich auch ihr eigenes Objekt gefunden, das ihr keine andere Wissenschaft streitig machen kann . . .“ (Troll 1950, 163). Das Forschungsobjekt ‚Landschaft‘ erlaubte die Integration von Mensch und Natur, d.h. die Verwissenschaftlichung des klassischen geographischen Paradigmas, und steht somit für die Einheit des Fachs. Bei Bobek und Schmithüsen kommt dies fast einem Glaubensbekenntnis gleich: „Belebte und unlebte Natur und Geist sind also in der geographischen Substanz verschmolzen. Darin liegt die innere Einheit der Geographie begründet und die Tatsache, daß diese weder Natur- noch Geisteswissenschaft allein sein kann“ (Bobek, Schmithüsen 1949, S. 113). Etwas sachlicher, aber im Kern identisch, äußert sich heute Goodchild über die disziplinäre Relevanz von GIS: „Of course, part of the attraction of GIS is its ability to surmount the human/physical divide“ (Goodchild 1995, S. 34).

GIS erfüllt damit gleich mehrere Funktionen: Es scheint (endlich!) die Möglichkeit zu eröffnen, die Landschaft (moderner: die ‚Umwelt‘) in ihrer unendlichen Komplexität in ganzheitlich-geographischer Manier zu analysieren. Zudem ermöglicht es die Zusammenführung von ‚quantitativer‘ und ‚angewandter‘ Geographie und die unreflektierte Wiedereinführung der quantitativen Forschungslogik. Die Anwendbarkeit und Relevanz von Verkehrs- und Wegwahlmodellen, der Modellierung von Schadstoffflüssen in terrestrischen Ökosystemen u.ä. ist leicht zu sehen und vermag so zu überzeugen, daß sich eine theoretische Diskussion gewissermaßen von selbst erledigt. Unter dem Deckmantel der angewandten Forschung werden alte forschungslogische Zweifel eskamotiert.

Auch Zweifler bestreiten weder die Anwendungsrelevanz noch den wissenschaftlichen Wert von GIS, sondern beklagen höchstens den Verlust einer heilen Welt: „Does society really want a Forest Service of GIS users at computer terminals rather than one of rangers on horseback?“ (Goodchild 1995, S. 48). Hier zeigt sich noch eine ganz andere (vielleicht die wichtigste, allerdings höchst widersprüchliche) Funktion von GIS: In dem Maße, in dem die Arbeit *mit* GIS die innere Einheit des Fachs vorantreibt (indem alle relevanten Fragen beantwortet werden, die die Menschheit bewegen), wird GIS selbst zum sinnstiftenden Integrationsobjekt, das mehr und mehr den geographisch-epistemologischen Platz der Landschaft besetzt. An die Stelle der von Hard herausgestellten „Apotheose der ‚Idee der Landschaft‘“ (Hard 1982a, S. 160) tritt jetzt die Apotheose der GIS-Idee, nach der die Landschaft als beforschte physisch-humane ‚Umwelt‘ zu retten, als Forschungsparadigma dagegen erneut zu zerstören ist. In diesem Sinne – und in einer weiteren Dekodierung des Akronyms – dient GIS in doppelter Hinsicht, sowohl als (technisches) Forschungswerkzeug als auch als (inhaltsleeres) Forschungsparadigma, als Geographisches Integrations-System.

## 1.4 GIS als Geometrisches Imaginations-System

Die Herleitung einer wissenschaftstheoretischen Grundlegung von GIS ist offensichtlich reichlich verwickelt. Zunächst wird die Anwendbarkeit von GIS als Werkzeug betont, das per se 'gut' ist, da es der Erhaltung der Lebensgrundlagen des Menschen dient. Nun mögen einzelne GIS-Anwendungen diesen Anspruch einlösen – als akademisches Forschungsprogramm ist ein solches Gütesiegel allein natürlich unbrauchbar. Die Akademisierung von GIS scheint jedoch zu gelingen, indem sich die GIS-Gemeinde klassischer geographischer Paradigmen bedient: Das mit GIS beforschte Objekt (die 'räumliche' physisch-humane 'Umwelt') ist zum einen erdräumlicher Natur und von geographisch relevanter Größe und Gestalt, zum anderen zeigt es sowohl landschaftlich-ganzheitliche als auch Mensch-Naturthematische Züge. Obwohl für ganzheitliche Erd(teil)raum-Betrachtungen traditionell die Geographie zuständig ist, kann das Fach seinen Anspruch heute, als Teil einer hochgradig ausdifferenzierten Wissenschaftslandschaft, allerdings nicht mehr allein erfüllen, und der holistische Charakter geographischer Forschungstechnik kann nur gerettet werden, indem sich die moderne GIS-Geographie als multisziplinäre Forschungsgemeinschaft versteht. Die Integration dieser Gemeinschaft gelingt zunächst über den "Raumbezug an sich" (Bartelme, a.a.O.), als angewandte Spatial Analysis, Geoinformatik oder Spatial Information Theory. Die erdoberflächliche Integration ist zwar noch nicht das eigentlich sinnstiftende Element der GIS-Gemeinde, sie ist jedoch notwendige Voraussetzung für die Schaffung eines interdisziplinären Interesses an der traditionellen geographischen Methodologie. Dieses kristallisiert schließlich im gemeinsamen Fokus auf eine 'raumbezogene' Technik, welche wiederum die notwendige Voraussetzung für die Einlösung des (selbst auferlegten) Gelübdes zur Erhaltung der menschlichen Lebensgrundlagen ist.

Der argumentative Kreis schließt sich über den Raum wie von selbst: GIS als Computerprogramm ist notwendig, weil mit seiner Hilfe eine Vielzahl praktischer Probleme auf interessante und fruchtbare Art und Weise bearbeitet werden kann (das soll hier nicht bestritten werden!). GIS als Konzept bietet Ansätze zur geographietheoretischen Betrachtung, weil die zu bearbeitenden Probleme ganzheitlich zu erfassen sind und daher im Sinne der klassisch geographischen Forschungslogik über die Erdoberfläche integriert werden müssen (und das können nun mal Geographen). GIS als Technik bietet schließlich selbst ein sinnstiftendes Forschungsparadigma, weil darüber 'raumbezogene Theorie' und praktische Arbeit unauflöslich verknüpft werden können. Der 'Raum' ermöglicht auf diese Weise, das praktische Tun als wissenschaftstheoretisch fundiertes Tun zu deuten.

Die Kompetenz der Geographen in Sachen 'Raum' ist unbestritten, und dennoch liegt genau im 'Räumlichen' der Schwachpunkt der Argumentation. Denn ähnlich der mehrfachen Bedeutung von GIS (als Hard- und Software, als gedankliches Konzept, als Forschungsstrategie, als Technik, als Werkzeug etc.) bietet auch der 'Raum' eine Menge von „bedeutungsverschiedenen Raumvokabeln“ (Hard 1986, S. 82, Fn. 1; vgl. auch Bartels 1974). Innerhalb der Geographie wird z.B. zwischen dem Erdraum als Container (z.B. Naturraum) und dem Erdraum als Ordnungsrelation (z.B. das Christallersche System der Zentralen Orte) unterschieden.<sup>10</sup> Daneben existieren eine Reihe weiterer Raumkonstrukte, und deren Anzahl steigt proportional zur Anzahl der Fachdisziplinen, die von der GIS-Gemeinde zu den 'Raum'wissenschaften gezählt und unter einem räumlichen Dach vereinigt werden. In der GIS-Gemeinde begegnet uns das 'Räumliche' als physisch materielle Umwelt, in der wir navigieren (Landau 1986, 1991; McNamara 1986; McNamara et al. 1989; Dieberger 1995), als Metapher für graphische Software-Oberflächen (Kuhn 1993), als kognitive Bildschemata (Mark 1989; Kuhn, Frank 1991), als Eigenschaften kognitiver Bildschemata (Kosslyn 1980, 1994), als logisches Kalkül (Mark, Frank 1990; Bittner, Frank 1997), als Element natürlicher Sprachen (Frank, Mark 1991; Landau, Jackendoff 1993;

<sup>10</sup>Insbesondere für sozialwissenschaftliche GIS-Anwendungen ist dieser zweite Aspekt von entscheidender Bedeutung, aber mit GIS nicht oder nur schwer zu realisieren (z.B. Anselin, Getis 1993).

Broschart 1995), als Grundlage hierarchischer Schlußfolgerungen (Hirtle, Jonides 1985; McNamara, Hardy, Hirtle 1989; Car, Frank 1994; Holdings 1994), als kinästhetische Erfahrung (Lakoff 1987; Johnson 1987), als Technik zur Erforschung mentaler Modelle (Byrne, Johnson-Laird 1989; Johnson-Laird, Byrne 1991), als Konstruktionshilfe mentaler Modelle (Glenberg, Langston 1992; Rauh, Schlieder, Knauff 1997), als Eigenschaft mentaler Modelle (Klauer 1998), als Ausdruck sozialer Hierarchien (Keating 1995), als Element sozialer Kommunikation (Klüter 1986, 1987, 1994), als kognitive Karten (Downs, Stea 1982; Portugali 1996), als Diagramme und externe didaktische Hilfen (Larkin, Simon 1987; Schnotz 1992; Stenning, Oberlander 1995), usw. usf.

Für die GIS-Gemeinde sind all diese 'Räumlichkeiten' relevant für das *Design* von GIS. Ungeachtet der Vielzahl möglicher Raumkonstruktionen steht jedoch das mit GIS zu behandelnde *Objekt* als ein „makroskopisch gesehen . . . irgendwie zusammengehöriges“ (Bobeck, a.a.O.) geographisches Wesen außer Frage. Dies ist insbesondere erstaunlich, da das geographische Wesen in der innerfachlichen Diskussion längst als Artefakt entlarvt worden ist. Raumbezogene Probleme und raumbezogene Daten existieren nun einmal nicht per se; Informationen, Problemstellungen, Daten etc. lassen sich nur (mehr oder weniger sinnvoll) räumlich projizieren. Eine räumliche Projektion dient in der Regel der Einführung eines Vergleichsmaßes bzw. 'objektiven Kodes', mit Hilfe dessen Interessen zwischen unterschiedlich vorinformierten Beteiligten einfacher kommunizierbar sind (vgl. Klüter 1986). Selbst ein Flächennutzungskonflikt, ein *raumplanerisches* Problem mit zutiefst geographischer Komponente, ist in diesem Sinne weder räumlich noch raumbezogen. Vielmehr geht es z.B. um ökonomische versus ökologische Interessen, die als solche unräumlich sind und erst durch eine kartographische Projektion (z.B. in Form eines Flächennutzungsplans) *verräumlicht* werden. Die Tatsache, daß es sich beim strittigen Objekt im Fall eines Flächennutzungskonflikts um eine zweidimensionale geometrische Grundform handelt, die erdoberflächlich verortbar ist, ist ebenso selbstverständlich wie bedeutungslos.

Daß räumliche Abbildungen in der Geographie besonders häufig auf Abbildungen der erdräumlichen Geometrie und damit auf Karten hinauslaufen, ist angesichts der Fachhistorie leicht einzusehen. Unabhängig davon bieten erdoberflächliche Projektionen sicherlich auch eine besonders leicht zugängliche, da mit der Alltagswelt unmittelbar vereinbare Form der räumlichen Abstraktion (vgl. z.B. Lakoff 1988). Aber auch in Bereichen, die der Geographie fern sind, spielen räumliche Ordnungsrelationen eine entscheidende Rolle für die Informationsvermittlung und Kommunikation. Man muß kein GIS-Experte sein, um die Vorteile zu erkennen, die ein dicker Pfeil auf dem Display des Fahrzeugnavigationsgerätes gegenüber der schriftlichen Anweisung 'jetzt rechts abbiegen!', ein Poincaré-Schnitt gegenüber Newtonschen Bewegungsgleichungen oder ein Histogramm gegenüber einer Häufigkeitstabelle hat. Fast jeder Versuch, komplizierte Sachverhalte zu vermitteln, läuft darauf hinaus, sie in Form von Zeichnungen oder Diagrammen und eben auch Karten zu veranschaulichen. Verräumlichungen und damit auch GIS-Anwendungen haben damit sowohl innerhalb als auch außerhalb der Geographie eine überwiegend didaktische Funktion.

Verräumlichungen im Sinne von Geometrisierungen entsprechen in ihrer einfachsten Form solchen Raumabstraktionen, die Klüter (1986) als Kulissen bezeichnet. Er illustriert diesen Begriff anhand eines Arbeiters, der eine ihm fremde Maschine bedienen soll (vgl. auch Hard 1986). Eine Einweisung in seine Tätigkeit könnte erfolgen, indem ihm die Funktion der Maschine als Teil des gesamten Produktionsablaufs erläutert würde, inklusive aller organisatorischen und technischen Details. Der Arbeiter hätte dann die Einsicht in die Notwendigkeit seiner Aufgabe und könnte diese mit vollem Verständnis erfüllen. Wenn das nicht realisierbar ist, könnte man ihm die relevanten Details der Maschine erläutern, die er bedienen soll. Am einfachsten wäre für alle Beteiligten jedoch die Erläuterung in Form eines geometrischen Kodes, z.B. in Form einer Skizze, auf der die notwendigen Bewegungsabläufe in chrono- und chorologischer Reihenfolge verzeichnet sind: Drücke zuerst den Knopf hier, dann gehe dort hinüber und ziehe diesen Hebel, tue soundso lange jenes und beginne dann von vorn.

Der 'Raum' fungiert für Klüter also als Element sozialer Kommunikation, indem räumliche (geometrische) Abbildungen es erlauben, komplexe und/oder abstrakte Informationen auf sehr effektive Weise durch räumliche Informationen zu substituieren. Der Erfolg von GIS ist somit keineswegs dadurch zu erklären, daß die meisten Fragestellungen, Probleme oder Interessenslagen in der „real world“ primär räumlicher Natur bzw. raumbezogen sind. Verantwortlich für den Erfolg ist vielmehr, daß mit GIS offenbar besonders anschauliche und leicht kommunizierbare Raumabstraktionen erzeugt werden können, die direkt an unsere alltagsweltlichen Erfahrungen eines uns umgebenden dreidimensionalen geographischen Raumes anknüpfen. Eine kartographische Abbildung liefert also eine hervorragende Metapher, indem sie eine Abbildung abstrakter Phänomene (Geld, ökologische Wertigkeit, ...) auf konkrete Phänomene (Flüsse, Straßen, Flächen, ...) ermöglicht. Doch der eigentliche 'Trick' Geographischer Informationssysteme besteht nicht in der Nutzung der Karten-Metapher allein, sondern auch und vor allem darin, daß die grundlegenden *methodischen* Operationen, die mit GIS implementiert sind (v.a. Verschneidung und Pufferung; vgl. Albrecht 1996), auf eben dieser Karten-Metapher beruhen. Um beispielsweise die landwirtschaftlichen Flächen zu bestimmen, auf denen der Bodentyp 'Schwarzerde' vorherrscht und auf denen gleichzeitig Zuckerrüben angepflanzt werden, werden mit GIS die Karten (Layer) der Bodentypen und der Bodennutzung 'verschnitten'. Diese GIS-Operation ist ungemein anschaulich und vergleichbar mit der Überlagerung zweier durchsichtiger Kartenfolien: Sind Schwarzerden in der Bodenkarte gelb und Rübenäcker auf der Nutzungskarte blau eingetragen, ist das Ergebnis als grüne Fläche unmittelbar abzulesen. Im Grunde geht es dabei jedoch um die Anwendung Boolescher Algebra in dem Sinne, daß die Menge derjenigen Punkte der (gekrümmten und meistens als euklidische Ebene operationalisierten) Erdoberfläche bestimmt werden soll, denen sowohl das Attribut 'Schwarzerde' als auch das Attribut 'Zuckerrüben' zugeordnet ist. Mit GIS wird die entsprechende kartographische Verräumlichung dieser abstrakten Operation erreicht, indem jedem Junktor (in diesem Fall 'und', aber auch 'oder', 'nicht' usf.) eine entsprechende erdräumliche Flächenoperation zugeordnet wird (in diesem Fall 'Schnittfläche', aber auch 'Gesamtfläche', 'alle anderen Flächen' usf.; vgl. auch Tomlin 1990; MacEachren 1995). Die Anschaulichkeit einer methodischen Nutzung der Karten-Metapher liegt unter anderem in der Isomorphie von Boolescher Operation und räumlichem 'Objekt' begründet. GIS bietet damit einen unmittelbaren Zugang zu abstrakten Formen der Datenanalyse.

Auch der Verfasser geht davon aus, daß GIS und 'Raum' in einer eindeutigen Beziehung zueinander stehen und daß 'Räumlichkeit' das zentrale Element einer jeden GIS-Anwendung ist. Allerdings wird der 'Raum' hier weder als erdoberflächlicher Container, noch als argumentative Klammer für eine Einheit der Geographie, der Geoinformatik oder der 'Raumwissenschaften' aufgefaßt, sondern als Substitutionsmedium im Sinne Klüters. Analog zum Vorschlag Klüters, Sozialgeographie als „Technologie von Transformations- und Übersetzungsstrategien zur Analyse und Rationalisierung von Raumabstraktionen“ (Klüter 1986, S. 168) zu verstehen, wird hier angeregt, GIS als eine Technik zur Transformation und Visualisierung geometrisch projizierbarer Problemstellungen und zur Rationalisierung und Erzeugung geometrischer Raumabstraktionen zu begreifen. Im Gegensatz zur Auffassung, GIS diene dem Rasonieren über raumbezogene Probleme, wird GIS hier als Möglichkeit verstanden, mit Hilfe räumlicher Abstraktionen über beliebige Probleme zu rasonieren. Das führt uns zu einer vierten und letzten Dekodierung des GIS-Akronyms als 'Geometrisches Imaginationssystem'.

Während Klüter die Rolle von Raumabstraktionen aus sozialwissenschaftlicher Perspektive erklärt, liegt ein Schwerpunkt der vorliegenden Arbeit auf der Frage, wie der 'Raum' als Substitutionsmedium aus kognitionspsychologischer Sicht 'funktioniert'. Warum kommt gerade räumlich-geometrischen Abstraktionen ein besonderer didaktischer Wert zu, und was unterscheidet sie von schriftlichen, lautsprachlichen oder anderen Formen der Abstraktion? Worin liegt die tiefere Einsicht der GIS-Werbefachleute, wenn sie behaupten: „Mapping the information stored in your corporate da-



tabase allows you to see patterns, relationships, and trends that can't be seen in table format“ (ESRI-Werbung, z.B. in GIS-Europe 1997/12, S. 51). Diesen Fragen wird in Kap. 2 nachgegangen.

Wir werden sehen, daß dem Layerkonzept als Menge planarer Subdivisionen eine Funktion zukommt, die weniger Folge erdräumlicher Notwendigkeiten, sondern eher Ausdruck einer menschlichen, 'raumbezogenen' Informationsverarbeitung ist. Planare Subdivisionen beliebiger Ebenen (auch der Erdoberfläche!) sind primär nicht deshalb notwendig, weil auf der Erdoberfläche nicht zwei Dinge an derselben Erdstelle existieren können, sondern weil sie einen unmittelbaren Zugang zum diagrammatischen Rasonieren bieten (Kap. 2.4). Bevor die Funktion dieser Art von Geometrisierung vollständig ausgenutzt werden kann, muß das Verfahren zur Verschneidung von Flächen jedoch so verallgemeinert werden, daß es die Bestimmung beliebiger Partitionierungen erlaubt (Kap. 3). Die Stärken (und Schwächen) der vorgeschlagenen Technik werden in Kap. 4 abschließend anhand von Beispielen diskutiert.

## Kapitel 2

# Zur Funktionsweise geometrischer Abstraktionen

Ausgangspunkt unserer Untersuchung ist die schlichte Tatsache, daß wir uns als menschliche Lebewesen in der Regel irgendwo auf der Erdoberfläche befinden, daß wir uns dort zwischen verschiedenen Standorten bewegen wollen und daß uns daher Informationen über die Anordnung und Lage dieser Standorte zur Verfügung stehen müssen. Die uns umgebenden physischen Objekte (Wohnhäuser, Fabriken, Universitäten, Flughäfen etc.) sind auf der Erdoberfläche verortet (wo sonst?), und wir müssen zwangsläufig auf der Erdoberfläche navigieren, um gezielt von einer Erdstelle zu einer anderen zu gelangen.<sup>1</sup> Die Kartierung der physischen Umwelt (d.h. primär die graphische Beschreibung der Lage und Verteilung von physischen Objekten auf der Erdoberfläche) als Grundlage der klassischen Geographie ist daher eine notwendige, bereits seit dem Paläolithikum tradierte Kommunikationstechnik (Stea, Blaut, Stephens 1996, S. 345). Oder anders: „Die Fähigkeit, räumlich zu denken, gehört zur überlebensnotwendigen Grundausstattung des Menschen“ (Rauh et al. 1998, S. 1; vgl. auch Ardila, A. 1993).

Die Navigation aufgrund von Karten setzt eine gewisse Abstraktionsfähigkeit der navigierenden Individuen voraus, da die graphische Beschreibung in die physische Umwelt ‘zurückübersetzt’ werden muß. Die Fähigkeit, sowohl zwischen Repräsentation (der Karte) und Repräsentiertem (der Erdoberfläche) eindeutig zu unterscheiden als auch eindeutige Beziehungen zwischen ihnen herzustellen (also gewissermaßen in Gedanken zwischen ihnen hin und her zu springen), bildet die Grundvoraussetzung einer jeden kartenbasierten Kommunikation. Im folgenden soll nun die These belegt werden, daß Orientierungs- und Navigationsfähigkeiten und -techniken nicht nur gezielte Ortsveränderungen auf der Erdoberfläche erlauben (obwohl sie sich primär aus diesem Anlaß entwickelt haben mögen), sondern eine wesentliche Grundlage für das menschliche Denken, Deduzieren und Kommunizieren bilden. Diesen gedanklichen Rahmen formulieren McNamara, Hardy und Hirtle sehr treffend: „Because many human experiences take place in space, many of our thoughts and activities rely on memories of spatial relations“ (McNamara, Hardy, Hirtle 1989, S. 211).

Es wird daher ein Bogen geschlagen von menschlichen Orientierungs- und Navigationsleistungen (Kap. 2.1) über mentale Repräsentationen (Kap. 2.2) und die Möglichkeiten, aufgrund solcher Repräsentationen Schlüsse zu ziehen (Kap. 2.3), hin zur Funktionsweise von Raumabstraktionen (Kap. 2.4). Mit diesem kognitionswissenschaftlichen Parforceritt soll gezeigt werden, daß die Funktionsweise von

---

<sup>1</sup> Auch standortgebundene Individuen benötigen eine zumindest rudimentäre Fähigkeit zur räumlichen Orientierung, z.B. um ohne hinzusehen die Kaffeetasse statt der Computermaus zu greifen. Ähnliches gilt z.B. auch für Fluggäste, die zwar keine Vorstellung von der geographischen Lage des Heimatflughafens und des Ferienziels benötigen, sich jedoch auf verschiedenen Flughäfen zurechtfinden müssen.

kartographischen Repräsentationen der von anderen (nicht-karto)graphischen Repräsentationen sehr ähnlich ist. Zumindest besteht Grund zu der Annahme, daß geometrische Abstraktionen in Form von Bildern, Diagrammen u.ä. auch für das Verständnis primär gänzlich ‘unräumlicher’ (also nicht auf der Erdoberfläche verorteter) Sachverhalte von Interesse sein können. Als Ziel kann damit auch formuliert werden, die Funktionsweise von Raumabstraktionen im Sinne Klüters (1986) besser verstehen zu lernen.

## 2.1 Navigation

### 2.1.1 Navigation aufgrund physischer Wahrnehmung

Erdraumbezogenes Wissen – auch über metrische Zusammenhänge – wird offensichtlich schon in sehr frühen Phasen der kindlichen Entwicklung mental repräsentiert. So sind Kleinkinder bereits im Alter zwischen 16 und 24 Monaten fähig, metrische Distanzen als Kodierschema zur Lokalisation von in einem Sandkasten versteckten Objekten zu nutzen (Huttenlocher, Newcombe, Sandberg 1994).

Navigationsleistungen sind insbesondere weitestgehend unabhängig von der Möglichkeit, die physische Umwelt, durch die navigiert wird, visuell zu erfahren. Dies konnten Landau, Spelke und Gleitman (1984) zeigen, indem sie Navigationsaufgaben mit blinden Kindern durchführten: In einem Raum mit vier Standorten ( $M = \text{Mother}$ ,  $B = \text{Basket}$ ,  $T = \text{Table}$ ,  $P = \text{Pillows}$ ) wurde die 31 Monate alte, von Geburt an blinde Kelly, jeweils ausgehend von einer identischen Position bei ihrer Mutter, zu den anderen drei Standorten hin und wieder zurückgeführt. Nachdem das Mädchen die Strecken  $M-B$ ,  $M-T$  und  $M-P$  also unter Führung zurückgelegt hatte, war sie anschließend in der Lage, von einem der drei anderen Startpunkte aus ( $B$ ,  $T$  oder  $P$ ) die drei verbliebenen, nicht eigenständig erfahrenen Strecken selbständig zu finden ( $B-T$ ,  $T-P$  und  $P-B$ ). Diese Leistung kann nicht auf Stimuli der einzelnen Standorte zurückgeführt werden (akustische oder andere Hinweise wurden nicht gegeben, bzw. nur dann, wenn Kelly das Ziel bereits so gut wie erreicht hatte). Auch ist es unmöglich, daß das Mädchen die Strecken rein aus der Erinnerung zurücklegte (die zurückgelegten Strecken waren für sie völlig neu). Die erfolgreiche Navigation war nach Ansicht der Autoren daher nur möglich, weil Kelly aufgrund der unter Führung zurückgelegten Wege über eine integrierte Repräsentation des räumlichen Gesamtbildes verfügte, und zwar inklusive sowohl der topologischen als auch der metrischen Eigenschaften. Nur eine solche Repräsentation würde es Kelly nämlich erlauben, die Navigationsaufgabe selbständig zu lösen (die Repräsentation muß gewissermaßen triangulierbar sein).

Interessanterweise zeigte Kelly, abweichend von der tatsächlichen Anordnung der Standorte, eine Vorliebe für rechte Winkel und deren Vielfache. Eine ähnliche Präferenz des Orthogonale wurde von Landau (1991) auch in Experimenten beobachtet, die die Untersuchung räumlich-geometrischer Identifikationsleistungen zum Ziel hatten. Sie stellte fest, daß von Geburt an blinde Kinder dreidimensionale geometrische Objekte bereits im Alter zwischen 18 und 36 Monaten mit großer Systematik untersuchten. Mit  $2\frac{1}{2}$  Jahren erkannten die Kinder geometrische Ähnlichkeiten dieser Objekte aufgrund ihrer rein haptischen Erfahrungen mit großer Genauigkeit. Bereits im Alter von drei Jahren konnten die Kinder auch Objekte wiedererkennen, die vertikal und/oder horizontal rotiert waren – dies allerdings nur dann, wenn die Rotationen in Sprüngen von Vielfachen des rechten Winkels erfolgten (wie z.B. ‘auf den Kopf gestellt’, ‘auf die Seite gelegt’ etc.). Ebenso waren sie in der Lage, die einzelnen Teile der Objekte in orthogonalen Begriffen, wie ‘oben’, ‘unten’, ‘hinten’ etc., zu benennen.

### 2.1.2 Navigation aufgrund symbolischer Wahrnehmung

Die Fähigkeit, bildhafte Vorlagen als Repräsentationen physisch-realer Objekte zu begreifen, entwickelt sich sprunghaft zwischen dem 24. und 30. Lebensmonat (DeLoache, Burns 1994). Schon im

Säuglingsalter ist der Mensch fähig, ein reales Objekt (z.B. eine Banane) auf einer bildlichen Repräsentation dieses Objekts (z.B. dem Photo einer Banane) wiederzuerkennen. Dieser Wiedererkennungseffekt beruht jedoch auf den rein bildbezogenen Funktionen des passiven visuell-neurologischen Apparates (vgl. Kap. 2.2) und bedeutet nicht, daß die Differenz zwischen Repräsentation und Repräsentiertem tatsächlich verstanden wurde (ungeachtet des Unterschieds kauen Säuglinge auf beidem mit gleichem Eifer). Das Verständnis dieser Differenz und damit die Fähigkeit, ein Bild als Informationsquelle zu betrachten, entwickelt sich erst ca. ab dem zweiten Lebensjahr und wird von den Autoren im wesentlichen auf die Bewußtheit körperlicher Erfahrungen zurückgeführt (das Photo schmeckt weniger gut und/oder bereitet nach dem Verzehr Bauchweh; vgl. auch Kap. 2.3.3; siehe aber auch Piaget, Inhelder 1967; Piaget 1954; Inhelder, Piaget 1958).

Bereits drei- bis vierjährige Kinder können mit Hilfe von Karten Objekte in einem Zimmer auffinden und gegebene Routen verfolgen (Rutland, Custance, Campbell 1993). Gerber und Kwan (1994) berichten von einer Gruppe zwölfjähriger Schüler, die sowohl mit Hilfe einer Straßenkarte als auch mit Hilfe einer textlichen Beschreibung (einer sequentiellen Auflistung von Straßennamen) in der Lage waren, eine Route in einem suburbanen Raum zu planen und anschließend erfolgreich zu navigieren. Die unterschiedlichen Grundlagen zur Planung der Route führten jedoch zu gänzlich unterschiedlichen Wahrnehmungen der durchlaufenen Umgebung.

Auch die Fähigkeit, von Kartenvorlagen, d.h. von räumlich angeordneten Symbolen auf die räumliche Struktur der physischen Umwelt zu abstrahieren, setzt nicht notwendig die visuelle Wahrnehmung der symbolischen Repräsentation voraus. Landau (1986) präsentierte von Geburt an blinden Kindern eine 'Landkarte' in Form eines Bretts mit aufgeklebten Blöcken, die die Lage von Objekten im Versuchszimmer repräsentierten. Nach einer kurzen mündlichen Einweisung und der Klärung des eigenen Standortes waren die Kinder nach dem Studium der Kartenvorlage tatsächlich in der Lage, die Objekte im Zimmer zu finden, und zwar auch dann, wenn die Kartenvorlage in (in Vielfachen des rechten Winkels) horizontal und/oder vertikal rotierten Positionen präsentiert worden war.

Auch wenn, wie Landau einräumt, derartige raumbezogene Aufgaben von blinden Kindern bzw. Erwachsenen weniger gut ausgeführt werden als von sehenden, demonstrieren ihre Ergebnisse, „that blindness need not deter the child's understanding of some basic, yet quite abstract, functions of spatial knowledge that seem to require both an appreciation of symbolic devices, and the capacity to construct unified representations of space“ (Landau 1986, S. 222). Die Fähigkeit, von sowohl direkt als auch symbolisch, visuell ebenso wie haptisch wahrnehmbaren raumbezogenen Erfahrungen auf die räumliche Struktur der physischen Umwelt zu abstrahieren, ist demnach bereits im frühen Kindesalter grundsätzlich vorhanden. Sie ist unabhängig von der Möglichkeit, die physische Umwelt bzw. ihre symbolische Repräsentation visuell wahrzunehmen; vielmehr scheint die Fähigkeit zu Navigationsleistungen dem Menschen als Erdbewohner bereits 'in die Wiege gelegt' zu sein.

### 2.1.3 Kognitive Karten

Die gesamtäumlichen Repräsentationen haben viel gemein mit dem, was in der kognitionswissenschaftlichen und geographischen Literatur unter dem Begriff der *kognitiven Karte* (*mental map*, *cognitive map*) verstanden wird. Der Begriff geht zurück auf Tolman (1948), der ihn benutzte, um das Verhalten von Ratten und, in Analogie, von Menschen in ihrer physischen Umwelt zu beschreiben. Seine Experimente zeigen, daß Ratten, die gelernt haben, eine Futterkiste über einen bestimmten Ausgang ihres Käfigs aufzusuchen, auf das Entfernen der Futterkiste reagieren, indem sie alternative Ausgänge in ähnlicher Richtung ausprobieren. Tolman vermutete, daß Lebewesen zu Navigationszwecken eine kartenähnliche mentale Repräsentation ihrer physischen Umwelt entwickeln, die in ähnlicher Weise wie eine kartographische Karte strukturiert ist und auf einer Euklidischen Metrik beruht.

In der heutigen Interpretation bezeichnet der Begriff nach Downs und Stea eine mentale Repräsentation der Lage und Verteilung von Phänomenen auf der Erdoberfläche (der sogenannten physischen oder räumlichen Umwelt bzw. dem „every day physical environment“; Gärling, Böök, Lindberg 1979, S. 200; zit.n. Golledge, Stimson 1997, S. 234): „Kognitives Kartieren ist ein abstrakter Begriff, welcher jene kognitiven oder geistigen Fähigkeiten umfaßt, die es uns ermöglichen, Informationen über die räumliche Umwelt zu sammeln, zu ordnen, zu speichern, abzurufen und zu verarbeiten“ (Downs, Stea 1982, S. 23; vgl. auch dies. 1973, S. 7; 1977; zusammenfassend Kitchin 1994; Nelson 1996; Golledge, Stimson 1997, S. 224ff.; Curiel, Radvansky 1998).

Die Eigenschaften kognitiver Karten sind, insbesondere was die Übereinstimmung von Eigenschaften der Repräsentation (der kognitiven Karte) und des Repräsentierten (der ‘räumlichen Umwelt’) angeht, gründlich untersucht worden. So wird vielfach angenommen, kognitive Karten seien hierarchisch strukturiert und böten somit die mentale Entsprechung eines geographischen Regionensystems. Ferguson und Hegarty konnten z.B. zeigen, daß Probanden, die ein *räumliches Layout* (eine Anordnung von Objekten auf einer Fläche in Form einer Karte oder einer ‘Klötzchenwelt’) gelernt haben, Routen aus der Erinnerung in der Regel anhand markanter Objekte beschreiben, die als *Referenzpunkte* (*landmarks* oder *anchors*, z.B. Kirchtürme, große Straßenkreuzungen und andere markante Bauwerke) dienen, um die (relative) Lage von anderen Objekten zu definieren. Die mentale Routenplanung erfolgt damit ähnlich wie die physische Navigation. Je komplizierter das räumliche Layout ist, desto mehr Referenzpunkte werden für die Strukturierung der kognitiven Karte benötigt; ab ca. fünf Referenzpunkten führt dies bei den meisten Probanden zu einer Regionenbildung um die Referenzpunkte herum und schließlich zu einer Strukturierung der mentalen Karten in Form eines hierarchischen Regionensystems (Ferguson, Hegarty 1994).

Die Annahme einer hierarchischen Struktur kognitiver Karten ist nicht unumstritten. Die mentale Repräsentation räumlicher Layouts wurde auch als propositionales Netzwerk (Byrne 1979) oder als analoge Repräsentation kontinuierlich variierender Eigenschaften (Thorndyke 1981) interpretiert. Hierarchische Ansätze werden insbesondere seitens der der Geographie nahen Forschung postuliert und haben sich inzwischen überwiegend gegen nicht-hierarchische Ansätze durchgesetzt. Es wird davon ausgegangen, daß verschiedene erdräumliche Regionen in Form eines hierarchischen Baums gespeichert werden, und daß das Detailwissen über diese Regionen mit abnehmender Hierarchiestufe zunimmt (zusammenfassend McNamara 1986). Die Regionen können durch physische Grenzen (Straßen, Flüsse, Mauern u.ä.), wahrgenommene Grenzen (z.B. Linien auf einer Karte) oder auch gänzlich subjektive Grenzziehungen definiert werden (Hirtle, Jonides 1985). *Partiell hierarchische* Theorieansätze erlauben schließlich auch die Kodierung von Relationen zwischen Orten in verschiedenen Regionen. Das ist zwar speichertechnisch redundant und wenig effektiv, da auf (fast) alle derartigen Relationen auch über den Hierarchiebaum zurückgeschlossen werden kann, erhöht jedoch die Geschwindigkeit und Präzision des Erinnerungsvorgangs. Die überwiegend konstanten Antwortzeiten von Probanden, die nach Objekten sowohl innerhalb derselben als auch aus verschiedenen Regionen befragt wurden, legen partielle Hierarchien zumindest nahe (Kuipers 1978; Stevens, Coupe 1978; Davis 1981; McDermott 1981; Lloyd, Cammack, Patton 1996).

Erste Ansätze von räumlich-hierarchischen Schlußfolgerungen sind bereits bei Kindern im Alter zwischen 16 und 24 Monaten zu beobachten (Huttenlocher, Newcombe, Sandberg 1994). Hinweise für die Evidenz hierarchischer Theorien geben auch McNamara, Hardy, Hirtle (1989). Wenn wir nach der relativen Lage von Reno (Nevada) und San Diego (Kalifornien) gefragt würden, so würden wir wahrscheinlich antworten, daß San Diego westlich von Reno liegt. Tatsächlich liegt Reno jedoch nordwestlich von San Diego. Die Autoren erklären diese Beobachtung mit der hierarchischen Struktur von Städten, Bundesstaaten und Ländern, die dem folgenden Syllogismus<sup>2</sup> zugrunde liegt:

<sup>2</sup>Als Syllogismus wird eine ‘Dreiheit’ von Aussagen bezeichnet, die aus zwei Prämissen und einer Konklusion besteht.

Kalifornien liegt westlich von Nevada.  
San Diego liegt in Kalifornien und Reno in Nevada.  
Daher liegt San Diego westlich von Reno.

Die Ungenauigkeit im Schluß bzgl. der räumlichen Beziehung zwischen San Diego und Reno liegt demnach darin begründet, daß der Schluß auf der großräumigeren Beziehung der Bundesstaaten beruht. Eine Ableitung der Beziehung zwischen den Städten basiert also auf der räumlichen Hierarchie zwischen den Städten und den übergeordneten Bundesstaaten, und zwar so, daß übergeordneten Relationen beim Schließen ein höheres Gewicht beigemessen wird. Der gleiche Effekt liegt nach Ansicht der Autoren auch den folgenden beiden (falschen) Schlußfolgerungen zugrunde:

New York liegt an der Ostküste Amerikas.  
Santiago de Chile liegt an der Westküste Amerikas.  
Daher liegt New York östlich von Santiago de Chile.

Der Pazifik liegt westlich des Atlantiks.  
Der Panamakanal verbindet Panama (Pazifik) mit Colón (Atlantik).  
Daher liegt Panama westlich von Colón.

In diesem Zusammenhang ist auch der Effekt des *spatial priming* von Interesse. McNamara (1986) ließ seine Probanden einen fiktiven Stadtplan lernen, auf dem sowohl die räumliche Lage von Objekten (Häusern, Straßen etc.) als auch die Namen dieser Objekte verzeichnet waren. Der Stadtplan war von den Probanden deutlich wahrnehmbar in Regionen unterteilt. Nach der Lernphase wurde den Probanden der gleiche Stadtplan vorgelegt, allerdings ohne Beschriftung, und sie wurden aufgefordert, die Namen bestimmter Objekte zu nennen. McNamara konnte zeigen, daß die Namen nicht gleichmäßig schnell erinnert wurden, vielmehr wurde der Name eines aktuell geprüften Objektes dann besonders schnell erinnert, wenn das zuvor geprüfte Objekt in der räumlichen Nähe des aktuellen Objekts lag. Zudem führten Objekte, die innerhalb derselben Region lagen, bei gleicher Distanz zu einem wesentlich deutlicheren Priming-Effekt als Objekte, die in verschiedenen Regionen lagen (vgl. auch McNamara, Ratcliff, McKoon 1984; McNamara et al. 1989; Sanocki, Epstein 1997; zusammenfassend zuletzt Curiel, Radvansky 1998). Holdings (1994) führt diesen Effekt auf die hierarchische Struktur des mental gespeicherten Wissens zurück: Je näher zwei Objekte (erdräumlich) zueinander liegen, desto näher werden sie auch in dem hierarchischen Regionenbaum gespeichert und desto weniger Hierarchiestufen müssen in diesem Baum auf dem 'Weg' von einem zum anderen Objekt durchlaufen werden. Spatial Priming ist sogar für solche kognitiven Karten nachzuweisen, die aufgrund von verbalen Beschreibungen konstruiert wurden (Denis, Goncalves, Memmi 1995; vgl. aber Langston, Kramer, Glenberg 1998).

Das mentale Regionensystem führt schließlich auch zu metrischen Fehleinschätzungen. Nach Wilton (1979) können Richtungsangaben zwischen erdräumlichen Regionen (hier: England und Schottland) schneller verifiziert werden als innerhalb von Regionen. Newcombe, Liben (1992) zeigen, daß Probanden, die ein räumliches Layout physisch durchlaufen haben, Luftliniendistanzen zwischen solchen Objekten überschätzen, die durch undurchsichtige Barrieren voneinander getrennt sind. Eine ebensolche Fehleinschätzung von Luftliniendistanzen berichtet McNamara (1986) bzgl. physischer Barrieren, die zwar durchsichtig sind, aber umgangen werden müssen. Partiiell eingeschränkte Möglichkeiten, eine physische Umwelt zu erfahren, führen demnach zur Organisation kognitiver Karten in Form von subjektiven Regionen, die die Fähigkeiten, Distanzen einzuschätzen oder Karten zu zeichnen entscheidend beeinflussen und deren geometrische Eigenschaften nicht notwendig mit denen der repräsentierten Verhältnisse auf der Erdoberfläche übereinstimmen (vgl. auch Hirtle, Jonides 1985, Ferguson, Hegarty 1994; Holdings 1994; Curiel, Radvansky 1998; Daniel, Denis 1998; Rothkegel, Wender, Schumacher 1998).

### 2.1.3.1 Kognitive Karten und Geographie

Die Definitionen und experimentellen Untersuchungen machen bereits deutlich, daß kognitives Kartieren offensichtlich als Spezialfall einer allgemeinen Gedächtnisleistung angesehen wird (vgl. Kap. 2.2), allerdings – und das wird häufig übersehen – *nicht* aufgrund von Eigenschaften des mentalen Prozesses selbst, sondern einzig aufgrund von Eigenschaften der mental verarbeiteten Phänomene: Eine Gedächtnisleistung wird dann (und nur dann) als kognitives Kartieren bezeichnet, wenn Wahrnehmungen oder Vorstellungen *erdräumlich* verorteter (oder sinnvoll verortbarer) Phänomene verarbeitet werden. Hier liegt die vielfach betonte Parallele zwischen Psychologie (Kognitionspsychologie, ‘Environmental Psychology’) und Geographie (‘Behavioral Geography’), deren Beachtung Lloyd bereits 1982 einforderte: „There is a clear need for geographers to focus more attention on theories concerning how cognitive information is internally stored and processed if we intend to understand how such information is used in spatial decision making“ (Lloyd 1982, S. 532). Eine „spatial decision“ ist natürlich erdräumlicher Art, d.h. es geht ausschließlich um Entscheidungen, die einen erdräumlichen Ausdruck finden<sup>3</sup> (vgl. auch Gärling, Golledge 1993; Walmsley, Lewis 1993; Kitchin 1993, 1997b; Golledge, Stimson 1997).

Vorstellungen, daß kognitive Karten wirkliche Karten sind (als Teil des Hippocampus, wie z.B. O’Keefe, Nadel 1978 und O’Keefe 1991 vorschlagen) oder als funktionale Analogie aufgefaßt werden sollten (wie bei Lloyd 1982 zu lesen ist), werden heute überwiegend abgelehnt. Vielmehr stellen sie ein hypothetisches Konstrukt dar, mit dessen Hilfe ‘räumliches’ Verhalten sowohl beschrieben werden kann als auch erklärt werden soll (z.B. Kitchin 1994a). Trotz dieser Abstraktheit sind kognitive Karten für die Geographie von besonderer Attraktivität, denn sie repräsentieren „information about environments that are either known to exist or are imagined but not necessarily present. Any given map, therefore, may be a mixture of information received at quite disparate time periods, and at any particular point in time may be incomplete, more or less schematized, or distorted, and may contain fictional or hypothetical information, or relics of the past which no longer exist“ (Golledge, Stimson 1997, S. 234; zu weiteren Definitionen vgl. ausführlich Kitchin 1994a). Mit anderen Worten: Eine kognitive Karte ist nicht mehr und nicht weniger als die mentale Repräsentation der klassischen geographisch-ganzheitlichen Welt, und zwar sowohl in ihrer räumlichen als auch in ihrer zeitlichen Dimension (vgl. Kap. 1.3).

Kognitive Karten enthalten nicht nur ‘räumliche’ (geometrische) Informationen, sondern auch attributive Werte in Form von Faktenwissen, Bildern, Vorstellungen, Bedeutungen u.v.m. Sie werden daher auch als Mengen von Wissensstrukturen unterschiedlicher Präzision gedeutet, die über Referenzpunkte geometrisch (mehr oder weniger korrekt) miteinander in Beziehung gesetzt werden können. Da liegt natürlich der Vorschlag nahe „... that the cognitive map should be viewed as an *internalized geographic information system* (GIS)“ (Golledge, Stimson 1997, S. 236; vgl. Golledge, Bell, Dougherty 1994; Hervorhebung im Original). Die Parallele ist eindeutig, denn sowohl mit GIS als auch mit kognitiven Karten werden Informationen symbolisiert und kodiert, werden geometrische und nicht-geometrische Informationen verknüpft, werden Teilmengen großer Wissensbestände über geometrische Eigenschaften selektiert etc., und „It is difficult to think of a single functionality embedded in the GIS that does not have a parallel in human-information processing capability“ (ebd.).

Der einzige Unterschied, den die Autoren sehen, „... is that in the GIS, once activated, a manipulative procedure should be carried out quickly and accurately, whereas in humans, although the ability to perform a manipulative activity may in principle be within their intelligence realm, there are many personal and societal barriers that may induce error and inhibit the use of such a skill“ (ebd. S.

<sup>3</sup>Der Leser möge sich an die Diskussion in den Kapiteln 1.2- 1.4 erinnern und sich fragen, ob menschliche Entscheidungen existieren können, die *keinen* erdräumlichen Ausdruck finden!

236). Während GIS also ein praktisches Werkzeug zur exakten Bearbeitung, Analyse und Speicherung 'raumbezogener' Informationen ist, sind die vergleichbaren menschlichen informationsverarbeitenden Prozesse ungenau und störanfällig (ebd. S. 236/237). Golledge und Stimson, die an anderer Stelle für GIS als Instrument zur Unterstützung und Analyse 'raumbezogener' Entscheidungen werben (ebd. S. 60f., S. 343f.), hebeln sich mit diesem Argument gewissermaßen selbst aus. Die Frage ist nämlich, wozu GIS bei der Analyse eines Spatial Behavior überhaupt eingesetzt werden kann, wenn die das Spatial Behavior beeinflussenden Faktoren („many personal and societal barriers“) individueller und sozialer Natur und eben gerade *nicht* exakt zu ver(erd)räumlichen sind (nicht als Elemente der kognitiven Karte und schon gar nicht mit GIS).

Jede Erklärung menschlichen Verhaltens mit Hilfe eines kognitiven Kartenkonstrukts übersieht zwangsläufig, daß Menschen sich nicht in bestimmter Weise verhalten, *weil* ihre kognitiven Karten das so vorgeben; vielmehr ist das, was als kognitive Karte bezeichnet wird, *Ausdruck* individueller Verhaltensmuster bzw. individuell angestrebter Verhaltensweisen. Kognitive Karten und insbesondere deren zeichnerischen Niederlegungen können daher maximal als Raumabstraktionen aufgefaßt werden, mit Hilfe derer über menschliches Verhalten (fachübergreifend) kommuniziert werden kann. In diesem Sinne ist Lloyd nur zuzustimmen, wenn er sagt: „Mental imagery research should be of particular value to geographers interested in using maps to communicate information“ (Lloyd 1982, S. 533). Im Unterschied zur geographischen Auffassung werden hier jedoch nicht ausschließlich erdraumbezogene (kartographische bzw. klassisch-geographische) Informationen betrachtet, die über Karten bzw. Bilder kommuniziert werden, sondern alle Informationen, die in sinnvoller Weise geometrisierbar sind. Versuchen wir daher im folgenden, den Weg in umgekehrter Richtung zu beschreiten, indem wir kognitive Karten (oder generell die „Fähigkeit, räumlich zu denken“; Rauh et al., a.a.O.) nicht als internalisiertes GIS auffassen, sondern GIS als deren (partielle) Externalisierung begreifen. Bevor wir diesen Gedanken weiter verfolgen, sollen jedoch zunächst einige grundlegende Erkenntnisse der Kognitionsforschung betrachtet werden.

## 2.2 Gedächtnis

Unter dem Begriff *Gedächtnis* fassen wir im weiteren jene Funktionen zusammen, die es uns ermöglichen, Wahrnehmungen, Erfahrungen und Erkenntnisse zu speichern, zu verarbeiten und später wieder zu vergegenwärtigen. Als *Wahrnehmung* bzw. *Perzeption* bezeichnen wir die Informationsaufnahme eines Individuums über eines oder mehrere seiner Sinnesorgane. *Kognition* bezeichnet schließlich die Erkenntnisgewinnung aufgrund wahrgenommener, erfahrener, gelernter oder erinnelter Informationen, d.h. einen Verarbeitungsprozeß, der auf gegebenen Informationen aufbaut (vgl. Baddeley 1997).

Mit den folgenden Ausführungen wird versucht, das heutige Verständnis der Funktionsweise des menschlichen Gedächtnisses in groben Zügen nachzuzeichnen. Es wird sich zeigen, daß die Entwicklung der Gedächtnisforschung im wesentlichen auf eine fortschreitende Untergliederung cerebraler Funktionen in verschiedene Systeme und Subsysteme hinausläuft. Wir sind dabei weniger an den neurowissenschaftlichen Grundlagen oder neurophysiologischen Details interessiert, sondern vielmehr an funktionalen Erklärungsansätzen, d.h. an Versuchen, das Gedächtnis und seine Leistungen (insbesondere die raumbezogenen) als Ausdruck von informationstragenden Strukturen zu verstehen, die auch ohne Kenntnis der Physis plausibel sind (neurowissenschaftliche Beiträge zur raum- und bildbezogenen Kognition finden sich z.B. in Olson, Bialystock 1983; Stiles-Davis, Kritchesvky, Bellugi 1988; Barlow, Blakemore, Weston-Smith 1990; Paillard 1991; Gazzaniga 1995; Kosslyn 1995a; Payne, Conrad 1996).



### 2.2.1 Kurz- und Langzeitspeicher

Die moderne Sichtweise des Gedächtnisses ist die eines funktional stark ausdifferenzierten Systems (vgl. im folgenden Baddeley 1997, S. 37ff.; auch Gathercole 1997), das keineswegs der bloßen Speicherung von Informationen dient („... the primary function of memory is, in fact, not memorizing“; Garnham 1997, S. 149). Erste Hinweise darauf gaben Experimente in den 60er Jahren. Postman und Phillips (1965) legten z.B. mehreren Versuchspersonen eine Liste unzusammenhängender Wörter vor, die diese auswendig lernen sollten, um anschließend möglichst viele Wörter aus dem Gedächtnis wiederzugeben. Die Autoren konnten zeigen, daß, wenn die Aufforderung zur Wiedergabe der gelernten Wörter direkt an die Lernphase anschloß, tendenziell die zuletzt gelernten Wörter richtig wiedergegeben wurden. Dieser sog. *recency effect* trat dagegen nicht auf, wenn zwischen der Lern- und Wiedergabephase eine Pause eingelegt wurde. Eine simple Interpretation dieser Beobachtung führte zu der Annahme, daß neu gelernte Informationen in einer Art temporärem und instabilem *Kurzzeitspeicher* (STS = short-term store), früher gelernte Informationen dagegen in einem stabileren *Langzeitspeicher* (LTS = long-term store) gehalten werden.

Andere Studien konnten die Annahme zweier Gedächtnis-Subsysteme bestätigen. Murdock (1965) konnte experimentell belegen, daß der Kurzzeitspeicher (auch als *primäres Gedächtnis* bezeichnet) nur eine eng begrenzte Speicherkapazität aber sehr kurze Zugriffszeiten, der Langzeitspeicher (auch *sekundäres Gedächtnis*) dagegen eine sehr große Speicherkapazität aber relativ lange Zugriffszeiten aufweist. Auch systematische Fehler beim Wiedergeben auswendig gelernter Informationen konnten mit Hilfe der beiden Subsysteme erklärt werden: Bei der Wiedergabe amerikanischer Telephonnummern (incl. Buchstaben) wurde beobachtet, daß falsche Wiedergaben von Buchstaben überzufällig häufig auf Verwechslungen mit phonetisch ähnlichen Buchstaben zurückzuführen sind (der Buchstabe ‘P’ wurde beispielsweise häufiger mit dem in der englischen Sprache phonetisch ähnlichen Buchstaben ‘V’ verwechselt als z.B. mit dem visuell ähnlicheren Buchstaben ‘R’; Conrad 1964). Umgekehrt hat der Grad der phonetischen Verwandtschaft von Buchstaben einen negativen Einfluß auf das Erinnerungsvermögen, z.B. ist ‘PGTVCD’ im Englischen auch sehr kurzfristig schwieriger zu erinnern als ‘RHXKWY’ (*phonological similarity effect*; Conrad, Hull 1964).

Ähnliche Beobachtungen wurden bei der Rekapitulation von Prosatexten gemacht. Probanden wurden verschiedene Textpassagen vorgelesen, wobei einige Sätze gelegentlich wiederholt wurden. Die Probanden sollten dann entscheiden, ob der wiederholte Satz mit dem zuerst vorgelesenen Satz exakt identisch war oder nicht. Die Modifikationen des wiederholten Satzes waren entweder syntaktischer oder semantischer Art. Wurde ein Satz im direkten Anschluß wiederholt, waren die Probanden in der Lage, Veränderungen sowohl des Satzbaus als auch der inhaltlichen Bedeutung des Satzes gleichermaßen zu erkennen. Lagen jedoch einige andere Sätze zwischen den Vergleichssätzen, nahm die Fähigkeit, syntaktische Modifikationen zu erkennen, deutlich ab, nicht dagegen die Fähigkeit, Bedeutungsveränderungen festzustellen (Sachs 1967). Diese und weitere Beobachtungen führten zu der Annahme, daß Informationen im Gedächtnis gewissermaßen doppelt kodiert werden (*dual coding*), und daß diese Kodierungen im Kurzzeitspeicher überwiegend phonetisch-syntaktischer, im Langzeitspeicher überwiegend semantischer Natur sind.

#### 2.2.1.1 Das modal model

Die Dichotomie des Kurz- und Langzeitspeichers galt Ende der 60er Jahre als beherrschendes Paradigma der Gedächtnisforschung. Ein für diese Phase als charakteristisch und einflußreich geltendes Modell wurde 1968 von Atkinson und Shiffrin vorgeschlagen (vgl. auch Baddeley 1997, S. 44). Mit ihrem sog. *modal model* wird angenommen, daß wahrgenommene Informationen zunächst parallel von einer Reihe von Zwischenspeichern (sensory buffer stores) verarbeitet werden. Diese Zwischenspei-

cher halten die Information für die Weiterverarbeitung im speicherbegrenzten Kurzzeitspeicher (STS), der wiederum den Langzeitspeicher (LTS) mit Informationen versorgt. Nach der Modellvorstellung gelangen also Informationen, die von den Sinnesorganen perzipiert werden, nur über den STS in den LTS. Zusätzlich zur reinen Informationsspeicherung werden dem Kurzzeitspeicher damit Kontrollfunktionen zugeschrieben. Die Grundidee dieses Modells besteht also in der Annahme, daß neben der Informationsspeicherung und -wiedergabe noch andere, im weiteren Sinne organisatorische Gedächtnisfunktionen existieren. Atkinson und Shiffrin verlegten diese organisatorische Komponente in den STS, der für sie als temporäres *Arbeitsgedächtnis* (*working memory*) zur Unterstützung zahlreicher kognitiver Funktionen und damit gewissermaßen als ‘Schleuse’ zwischen Wahrnehmung und Speicherung fungiert. Die Wahrnehmung neuer Informationen führt gegebenenfalls zur Verschiebung des aktuellen Inhalts des STS in den LTS; werden keine Informationen aufgenommen, wird der STS nach einer Weile in den LTS entleert.

Trotz seiner sehr formalen Notation bietet das Modell von Atkinson und Shiffrin eine einfache Erklärung für den recency effect, der danach Ausdruck jener Informationen ist, die aktuell im Kurzzeitspeicher gehalten werden. Mit der Annahme, der STS sei gleichzeitig Speicher- und Kontrollmedium, prognostiziert das Modell jedoch auch, daß Menschen mit einem schwachen Kurzzeitgedächtnis zwangsläufig Probleme mit dem langfristigen Lernen haben (sie müßten weniger bzw. langsamer lernen). Diese Vorhersage konnte nicht bestätigt werden (Shallice, Warrington 1970). Desweiteren besagt das Modell, die Wahrscheinlichkeit, daß eine Information in den LTS verschoben wird, sei eine direkte Funktion ihrer Verweildauer im STS. Auch für diese Modellvorhersage konnten keine übereinstimmenden experimentellen Ergebnisse gefunden werden (z.B. Tulving 1966; Morton 1967; Nickerson, Adams 1979; zusammenfassend Baddeley 1997, S. 48).

### 2.2.1.2 Zweitaufgabentechnik

Trotz zahlreicher, offensichtlich falscher Vorhersagen, die das modal model trifft, blieb die Idee des Arbeitsgedächtnisses im Zentrum weiterer Untersuchungen. Mit diesem Begriff werden Gedächtnisfunktionen zusammengefaßt, die für (kurzfristige) kognitive Leistungen verantwortlich sind: „Working memory . . . is the term used to describe the alliance of temporary memory systems that play a crucial role in many cognitive tasks such as reasoning, learning and understanding“ (Baddeley 1997, S. 6). Die Frage war (und ist), wo das Arbeitsgedächtnis zu verorten ist, und zwar weniger im geometrischen Sinne (als Teil des Gehirns), sondern im Sinne einer Systematik der gedächtnisrelevanten Funktionssysteme. Es galt daher u.a., die Annahme von Atkinson und Shiffrin zu überprüfen, daß die Funktion des Arbeitsgedächtnisses dem Kurzzeitspeicher (STS) zuzuschreiben sei. Dafür bediente man sich eines Standardverfahrens der experimentellen Psychologie, der sog. *Zweitaufgabentechnik* (*dual-task technique*): Probanden werden aufgefordert, einfache Lernaufgaben zu bewältigen, mit denen ein Großteil der Kapazität des Arbeitsgedächtnisses absorbiert wird, während sie gleichzeitig andere Aufgaben erfüllen sollen, von denen ebenfalls angenommen wird, daß sie überwiegend Ressourcen des Arbeitsgedächtnisses in Anspruch nehmen. Ist diese Annahme korrekt, dann müßten beide Aufgaben um die begrenzten Ressourcen des Arbeitsgedächtnisses ‘konkurrieren’, d.h. es müßten Interferenzen zwischen den verschiedenen informationsverarbeitenden Prozessen (bzw. deren Resultaten) beobachtbar sein.

Obwohl zahlreiche unterschiedliche Vorstellungen über die Funktionsweise des Kurzzeitspeichers existieren, wird übereinstimmend davon ausgegangen, daß der STS eine eng begrenzte Kapazität hat, und daß diese Kapazität durch einfache und direkte Wiederholungsaufgaben in Anspruch genommen wird. Als einfacher, den STS beanspruchender Aufgabentyp eines Zweitaufgabenexperiments wird daher häufig eine Form der ‘memory span’-Technik gewählt. Nach Baddeley (1997, S. 29/30) ist der Erfinder dieser Technik der Londoner Schulmeister Joseph Jacobs, der Ende des letzten Jahrhunderts

eine Methode entwickelte, mit der er die mentalen Fähigkeiten seiner Schüler quantitativ zu vergleichen versuchte. Probanden werden dabei Symbolreihen (in der Regel Ziffernfolgen) vorgelegt, die sie mündlich und aus dem Gedächtnis wiederholen sollen. Typischerweise wird mit kurzen Reihen begonnen, die im Laufe des Experiments bis zu dem Punkt immer weiter verlängert werden, an dem der Proband sie nicht mehr fehlerfrei rekapitulieren kann. Die Länge der Reihen, die ein Proband zu 50% richtig wiederholen kann, wird als sein *memory span* bezeichnet. Führt man dieses Experiment mit Ziffernfolgen durch, spricht man auch vom *digit span*, der normalerweise zwischen sechs und sieben liegt. Die andere Aufgabe eines Zweitaufgabenexperiments besteht in der Regel in anspruchsvolleren Tätigkeiten, die im engeren Sinne kognitive Fähigkeiten erfordern, wie z.B. das Treffen von Schlußfolgerungen oder das Verständnis komplexerer Zusammenhänge. Baddeley (1968) gab z.B. Sätze vor, die eine Anordnung von Buchstaben sowohl beschreiben als auch zeigen. Die Aufgabe war, zu entscheiden, ob die Anordnung der Buchstaben mit der Beschreibung übereinstimmt. Die Sätze variierten dabei zwischen einfachen Aussagen, wie z.B.

$A$  folgt  $B$   $BA$  (wahr)

und komplexeren Aussagen, wie z.B.

Der Vorgänger von  $B$  ist nicht  $A$   $AB$  (falsch)

Wenn nun die Annahme korrekt ist, daß dem STS neben der Speicherfunktion auch die Funktion eines koordinierenden Arbeitsgedächtnisses unterliegt, dann müßten die kognitiven Leistungen bei der Lösung dieser Aufgaben durch die gleichzeitige Ausführung von digit span-Aufgaben erheblich beeinträchtigt werden. Geht man davon aus, daß die Kapazität des STS beim Erreichen des digit spans vollständig von der Wiederholungsaufgabe in Anspruch genommen wird, müßte an diesem Punkt die Ausführung einer weiteren kognitiven Leistung vollends unmöglich sein. Tatsächlich zeigte sich im Experiment der Effekt, daß sich die Antwortzeiten für die Schlußfolgerungen mit der Länge der wiederzugebenden Zahlenreihen verlängerten. Jedoch – und das ist für die Weiterentwicklung der Arbeitsgedächtnis-Hypothese von entscheidender Bedeutung – verlängerten sich die Antwortzeiten gegenüber denen ohne konkurrierende Wiederholungsaufgabe selbst bei Ziffernfolgen mit acht Ziffern (also beim Erreichen des digit spans) um nur 35%. Beinahe noch erstaunlicher ist die Beobachtung, daß selbst an diesem Punkt falsche Antworten nicht häufiger waren als bei geringeren Ziffernbelastungen (Baddeley 1968; vgl. auch Baddeley 1997, S. 50/51). Was aus diesen und ähnlichen Ergebnissen (z.B. Baddeley, Hitch 1974) folgt, ist klar: Der Kurzzeitspeicher ist kein unitäres System, vielmehr stellt das Arbeitsgedächtnis als funktionale Basis für kognitive Fähigkeiten ein eigenständiges Subsystem dar.

### 2.2.2 Arbeitsgedächtnis

Baddeley (1986 und 1997) schlägt daher ein Modell vor, nach dem das Arbeitsgedächtnis – als Teil des STS – aus einem zentralen Kontrollsystem und zahlreichen weiteren Subsystemen besteht. Insbesondere unterscheidet er zwischen zwei Subsystemen, die zum einen phonologische, zum anderen visuell-räumliche Informationen verarbeiten: Die sog. *artikulatorische Schleife* (PL = phonological bzw. articulatory loop) ist verantwortlich für die Speicherung und Manipulation einer geringen Menge sprachlicher Informationen, und der sog. *visuell-räumliche Notizblock* (VSSP = visual-spatial scratchpad; vgl. ebd.) leistet dasselbe für eine geringe Menge räumlicher und bildlicher Informationen. Das Kontrollsystem, die sog. *zentralen Exekutive* (central executive; vgl. ebd.), beschreibt Baddeley als eine Art ‘Organisator’, der Informationen auf die Subsysteme verteilt und von dort integriert (vgl. Baddeley 1997, S. 52ff.; Klauer 1998, S. 11).

### 2.2.2.1 Die artikulatorische Schleife

Mit der artikulatorischen Schleife werden nach Baddeleys Modellvorstellung Informationen im Sinne des dual codings (vgl. S. 25) grundsätzlich phonologisch kodiert (vgl. im folgenden Baddeley 1997, S. 52-69 sowie die dort angegebene Literatur). Sie besteht aus einem phonologischen Speicher, der sprachbasierte Informationen für ca. eineinhalb bis zwei Sekunden puffern kann, und einem Kontrollprozeß, der z.B. schriftliche in phonologische Informationen transformiert und diese im Speicher registriert oder von dort liest.

Aufgrund dieser einfachen Vorstellung können zahlreiche experimentelle Phänomene sinnvoll interpretiert werden. Der Effekt der phonologischen Ähnlichkeit von Informationen auf die individuelle Lern- bzw. Wiedergabefähigkeit (s.o.) läßt sich beispielsweise als direkte Folge der phonologischen Kodierung erklären: Phonetisch ähnliche Informationen werden ähnlich kodiert und sind daher beim Auslesen des phonologischen Speichers schwerer zu unterscheiden. Auf die Evidenz einer phonologischen Kodierung deuten Interferenzen zwischen visuellen Informationen und irrelevanten sprachlichen Nebengeräuschen hin (der sog. *unattended speech effect*). Colle, Welsh (1976) führten z.B. ein digit span-Experiment durch, in dem Probanden versuchten, visuell präsentierte Ziffernfolgen wiederzugeben. In einer Variante des Experiments wurde die Wiederholungsaufgabe von jemandem begleitet, der laut Texte in deutscher Sprache vorlas. Obwohl die Probanden diese Sprache nicht beherrschten, zeigte sich ein deutlicher Rückgang ihres digit spans. Trotz der Irrelevanz für die eigentliche Aufgabenstellung, visuell perzipierte Ziffernfolgen wiederzugeben, beansprucht der vokale Hintergrund offensichtlich dieselben Ressourcen des Arbeitsgedächtnisses.

Eine deutliche Bestätigung erfährt diese Interpretation durch Experimente, in denen anstatt eines unverständlichen Textes Zahlen vorgelesen werden. Salamé, Baddeley (1982; vgl. auch Salamé, Baddeley 1987 sowie Baddeley 1997, S. 53) wählten als Hintergrund sowohl konkrete Zahlwörter als auch Silben, die den Zahlwörtern phonetisch entsprechen (z.B. 'tun', 'woo' anstatt von 'one', two') sowie zweisilbige, phonetisch unähnliche Wörter (z.B. 'happy', 'tipple'). Die zweisilbigen Wörter verursachten gegenüber einer ungestörten Versuchsvariante einen leichten Rückgang des digit spans der Probanden. Die konkreten Zahlwörter wie auch die phonetisch ähnlichen Nonsense-Silben beeinflussten die Probanden dagegen wesentlich stärker. Diese beiden Einflüsse untereinander unterschieden sich jedoch, was den Grad der Störung betrifft, nicht. Der Befund spricht für die Modellannahme eines artikulatorischen Subsystems, das Informationen auf phonologischer Basis (und nicht auf Basis der Wörter oder ihrer Semantik) kodiert. Das Arbeitsgedächtnis wird damit als ein (zumindest partiell) sprachbasiertes System charakterisiert.

### 2.2.2.2 Der räumlich-visuelle Notizblock

Der räumlich-visuelle Notizblock (VSSP) als zweites wichtiges Subsystem in Baddeleys Modell des Arbeitsgedächtnisses ist verantwortlich für Gedächtnisleistungen, die wir, mehr oder weniger umgangssprachlich, als 'bildliche Vorstellungen' bezeichnen (vgl. Kap. 2.2.3). Auch der räumlich-visuelle Notizblock besteht in der Modellannahme aus einem passiven visuellen Speicher, der die bildhaften Informationen enthält, und einem aktiven Kontroll- und Wiedergabemechanismus, der abstrakte, räumlich-geometrische Informationen speichert und manipuliert (vgl. im folgenden Baddeley 1997, S. 71ff.; auch Klauer 1998, S. 16 sowie die dort angegebene Literatur).

Die Annahme eines räumlich-visuellen Notizblocks als separates Subsystem des Arbeitsgedächtnisses läßt sich, ähnlich wie die Annahme der artikulatorischen Schleife, durch Zweitaufgabenexperimente erhärten: Unsere Vorstellungskraft als Determinante unserer Lernfähigkeit und damit als Ausdruck eines Systems, das Informationen auf räumlich-visueller Basis kodiert, müßte sich durch räumlich-visuelle Zweitaufgaben negativ beeinflussen lassen. In der Tat weist z.B. Logie (1986) –

in Analogie zum unattended speech effect – einen *unattended picture effect* nach. Seine Probanden lernten Wortlisten, während sie sich u.a. auf einen Bildschirm konzentrieren mußten, auf dem in regelmäßigen Intervallen farbige Muster erschienen. Bei den Probanden, die aufgefordert waren, sich während des Lernens Bilder auszudenken, zeigte sich gegenüber denen, die ohne mentale Vorstellungen lernen sollten, ein deutlich negativer Einfluß der visuellen Zweitaufgabe auf ihre Wiedergabefähigkeit. In weiteren Experimenten erwies sich das bildhafte Lernen als robust gegenüber irrelevanten sprachlichen, das nicht-bildhafte Lernen als robust gegenüber irrelevanten visuellen Einflüssen.

Logies Resultate stellen die visuelle Komponente des räumlich-visuellen Notizblocks in den Vordergrund, während andere Autoren mit Hilfe ähnlicher Experimente oder aufgrund neurophysiologischer und – psychologischer Beobachtungen die räumlich-geometrische Komponente betonen. Farah et al. (1988a) beschreiben zwei separate Subsysteme für die Transformation und Repräsentation zum einen visueller, zum anderen metrischer Informationen. Kosslyn et al. (1989) unterscheiden abweichend – und nach Ansicht zahlreicher Autoren auch erfolgversprechender (zuletzt Cooper 1995) – zwischen kategorialen und metrischen räumlichen Repräsentationen. Danach sind kategoriale Repräsentationen solche, die sich auf generalisierte, abstrakte Eigenschaften einer räumlichen Struktur beziehen, die also im wesentlichen topologische Informationen beinhalten, z.B. ‘innerhalb von’, ‘unterhalb’ etc. Metrische Repräsentationen spezifizieren dagegen präzise Informationen über Koordinaten, Distanzen oder Winkel von und zwischen räumlichen Strukturen und deren Teilen. Diese Unterscheidung erscheint plausibel sowohl für die Fähigkeit zur Objekterkennung, d.h. für Transformationen visuell perzipierter Informationen, als auch für die Fähigkeit zu bildhaften Vorstellungen, d.h. für Transformationen sowohl visuell als auch nicht-visuell perzipierter Informationen (vgl. Kap. 2.2.3).

Übereinstimmend wird in jedem Fall von der Existenz eines – evtl. weiter untergliederten – visuell-räumlichen Subsystems des Arbeitsgedächtnisses ausgegangen, das (visuell und nicht-visuell perzipierte) Informationen in mentale Bilder transformiert und diese manipulieren kann. Ein weiteres Charakteristikum des Arbeitsgedächtnisses ist damit die (zumindest partiell) visuell-räumliche Verarbeitung von Informationen (zusammenfassend Cooper 1995; insbesondere über neuropsychologische Erkenntnisse berichten Kosslyn 1987; Farah 1988b; Farah et al. 1988b; Morrow, Ratcliff 1988; Gazzaniga 1995; Kosslyn 1995a; Tversky 1996).

### 2.2.3 Bilder und Vorstellungen

Eine wichtige Rolle bei der Klärung der Struktur von Gedächtnisleistungen spielt schließlich auch die Frage, inwieweit es funktionale Äquivalenzen zwischen unserer Wahrnehmung von physischen Bildern und unseren mentalen Leistungen aufgrund von Vorstellungen u.ä. gibt. Der englische Ausdruck ‘imagery’ und auch das deutsche ‘sich ein Bild von etwas machen’ deuten zumindest auf Parallelen zwischen Bildern und Vorstellungen hin.

Diese *imagery debate* führte in den 1970er Jahren zu zahlreichen experimentellen Untersuchungen der Eigenschaften mentaler Repräsentationen aufgrund von Bildvorlagen (zusammenfassend Lloyd 1982). Shepard und Metzler (1971) präsentierten ihren Probanden Zeichnungen zweier dreidimensionaler Objekte. Die beiden dargestellten Objekte waren jeweils unterschiedlich rotiert, ansonsten aber entweder identisch oder spiegelverkehrt dargestellt. Die Probanden sollten so schnell wie möglich entscheiden, welche der beiden Möglichkeiten vorlag. Shepard und Metzler beobachteten, daß die Zeit, die die Probanden für ihre Antwort benötigten, linear abhängig war vom Winkel, in dem die beiden Objekte zueinander standen – gerade so, als würden die Probanden tatsächlich eines der Objekte solange in Gedanken rotieren, bis es im selben Winkel wie das andere Objekt liegt, um dann ihre Entscheidung zu treffen. Sie folgerten, daß visuelle Vorstellungen auf einem analogen, d.h. stetigen Medium beruhen, das graduelle Manipulationen des mentalen Bildes erlaubt (vgl. auch Cooper, Podgorny 1976).

Die Annahme analoger Repräsentationen wird auch von Kosslyn (1980) bestätigt. Seinen Probanden wurde die Zeichnung eines Bootes vorgelegt. Nach genauer Ansicht der Zeichnung wurde diese entfernt, und den Probanden wurden Fragen über Details des Bootes gestellt. Folgte einer Frage über ein Detail am Bug des Bootes beispielsweise eine weitere bzgl. des Ruders im Heck des Bootes, benötigten die Probanden vergleichsweise mehr Zeit für die Antwort als bei einer Anschlußfrage, die sich auf die Ankerwinde in direkter Nähe des Bugs bezog. Die Probanden schienen das mentale Bild (die vorgestellte Zeichnung) gewissermaßen physisch zu scannen, wobei längere Scans auch mehr Zeit benötigten. Kosslyn konnte nachweisen, daß die für das Scannen von einem Teil des Bootes zu einem anderen benötigte Zeit eine (stetige) Funktion der Distanz zwischen diesen Teilen ist. Den gleichen Effekt konnte Kosslyn auch für solche Vorstellungen nachweisen, die die Probanden eigenständig, d.h. ohne Vorlage eines Bildes von vertrauten Objekten entwickelten.

Eine propositionale, d.h. diskrete Interpretation vorgestellter Bilder gibt dagegen Pylyshyn (1973, 1979; vgl. auch Baddeley 1997, S. 72-73), der die Ergebnisse der Rotationsexperimente durch eine Serie diskreter mentaler 'Sprünge' zu erklären versucht. Bildhafte Vorstellungen sind danach irrelevante Begleiterscheinungen eines Sets von Prozessen, mit denen raumbezogene, in Form diskreter Propositionen im Langzeitspeicher gehaltene Informationen mental verarbeitet werden. Insbesondere die Kosslyn-Effekte seien ebensogut mit Hilfe eines propositionalen Modells erklärbar, und zwar als Folge eines allgemeinen Erfahrungswissens ('tacit knowledge'; Pylyshyn 1981). Fragen über ein zeichnerisch dargestelltes Objekt würden die Probanden veranlassen, eine interne Simulation des dargestellten Objekts (nicht der Zeichnung dieses Objekts!) aufzubauen. Aufgrund der jahrelangen Erfahrung mit der visuellen Wahrnehmung der physischen Umwelt würden in einer solchen Simulation die zentralen Aspekte visuell perzipierter Erfahrungen reproduziert, was sich dann auf die Antwortzeiten auswirkte.

Obwohl die Auseinandersetzung um analoge und propositionale Repräsentationen eher informationstheoretische Aspekte berührt, ist sie doch primär durch die Frage motiviert, welche Ähnlichkeiten die mentalen Prozesse des 'Sehens' und des 'Sich-Vorstellens' aufweisen. Die Ergebnisse der Rotations- und Scanningexperimente deuten zumindest auf funktionale Parallelen hin: Die mentale Bearbeitung von bildhaften Vorlagen und von Vorstellungen (physischer Objekte) ist, was die relativen Bearbeitungsgeschwindigkeiten angeht, identisch. Es liegt daher nahe, die physische Durchführung der Objektrotation (das Drehen eines Gegenstandes in den Händen) sowie der Bildbetrachtung (das Abtasten eines Bildes mit den Augen) als Analogie für die entsprechenden mentalen Prozesse zu begreifen. Eine genauere Bestimmung der Natur dieser Analogie ermöglichten u.a. Zweitaufgabenexperimente von Farah (1985): Die Geschwindigkeit des visuellen Erkennens eines bestimmten Buchstabens wird durch die vorherige Aufforderung, sich diesen Buchstaben vorzustellen, deutlich erhöht (die Zeitspanne zwischen Aufforderung und Präsentation muß natürlich kleiner sein, als die zeitliche Speicherkapazität des visuellen Kurzzeitspeichers). Die Aufforderung, sich einen anderen Buchstaben vorzustellen als den, der anschließend visuell präsentiert wird, wirkt dagegen negativ auf die Lösungsgeschwindigkeit der visuellen Aufgabe. Offensichtlich besteht die funktionale Äquivalenz zwischen Bildern und Vorstellungen u.a. darin, daß beide die gleichen Ressourcen des Arbeitsgedächtnisses beanspruchen.

Diese These fand auch Bestätigung durch neuropsychologische Erkenntnisse. Nach Untersuchungen an Patienten mit partiellen Hirnverletzungen oder Messungen ereignisbezogener cerebraler elektrischer Potentiale sowie regionaler cerebraler Blutflüsse lassen sich beispielsweise visuelle Wahrnehmungsleistungen und Vorstellungsleistungen in denselben Gehirnzentren, wahrscheinlich sogar in denselben Neuronen, verorten (vgl. zusammenfassend Cooper 1995, S. 1578 sowie die dort angegebene Literatur). Interessant sind in diesem Zusammenhang schließlich anatomische Erkenntnisse, nach denen den cerebralen Hemisphären jeweils unterschiedliche Funktionen zugeschrieben werden können: Obwohl eine strikte Dichotomie schwer nachweisbar ist, besteht wenig Zweifel daran, daß die

linke Hemisphäre bei der überwiegenden Mehrheit der Menschen (ca. 95%) auf das Sprachverständnis spezialisiert ist, die rechte Hemisphäre dagegen überwiegend mit Navigationsaufgaben belastet wird (Millar 1994, S. 56f.). Kosslyn (1987; Kosslyn et al. 1989) lokalisiert nun die visuell-räumlichen Funktionssysteme in *beiden* cerebralen Hemisphären: Die linke Hemisphäre ('Sprache') ist verantwortlich für die Verarbeitung kategorialer räumlicher Informationen, wogegen die rechte Hemisphäre ('Navigation') auf die Verarbeitung räumlich-metrischer Informationen spezialisiert ist. Diese Resultate konnten in zahlreichen Experimenten bestätigt werden (z.B. Farah 1988b; Hellige, Michimata 1989; vgl. auch Rybash, Hoyer 1992; Hickok et al. 1996).

Diese Ergebnisse korrespondieren schließlich gut mit der inzwischen weithin akzeptierten *dual coding Theorie*. Paivio (1969, 1971, 1986) konnte mit seinen Untersuchungen des *assoziativen Lernens* zeigen, daß die individuelle Fähigkeit, Wörter zu erlernen und zu erinnern, mit den Angaben seiner Probanden, sich von den Wörtern 'ein Bild machen zu können', positiv korreliert. Bugelski, Kidd und Segmen (1968) entdeckten den verwandten Effekt, daß die Fähigkeit ihrer Probanden, Wortpaare zu assoziieren, zunahm, wenn sie aufgefordert wurden, sich für die Wörter während des Lernens ein Bild auszudenken, in dem die Wörter als bildliche Objekte miteinander interagieren. Dieses *dual coding* über 'bildhafte Eselsbrücken' war unabhängig davon, ob die Worte von vorneherein bildhaft waren, d.h. in einem ontologischen Sinn prinzipiell visuell wahrnehmbare Objekte bezeichneten, oder nicht. Bereits in diesen Arbeiten wurde deutlich, daß die bildliche Vorstellungskraft zum einen eine bestimmende Größe für Gedächtnisleistungen ist, und daß sie zum anderen nicht ausschließlich auf Dinge beschränkt ist, die visuell perzipiert werden können. Vielmehr scheint ihr eine generelle mediative Funktion für das Lernen und Erinnern zuzukommen.

#### 2.2.4 Kognitive Karten und Bilder

Die Verarbeitung bildhafter Vorstellungen (besser: Vorstellungen von Bildern) zeigt Übereinstimmungen mit tatsächlichen visuellen und haptischen Prozessen, da z.B. die für das physische Scannen von Bildern bzw. die Rotation von Objekten benötigte Zeit im hohen Maße positiv mit der Dauer mentaler Scan- und Rotationsoperationen korreliert. Der Mensch ist offensichtlich in der Lage, seine Wahrnehmungen mental so zu strukturieren, daß physische Operationen auch 'im Geiste' ausgeführt werden können, und verfügt damit über grundlegende Fähigkeiten, physische Aktionen vor ihrer Ausführung zu planen. Die Entwicklung mentaler Repräsentationen in Form von Bildern setzt dabei nicht die visuelle Wahrnehmung des zu repräsentierenden Objektes voraus, sondern funktioniert ebensogut aufgrund verbaler Beschreibungen: „... individuals are not only able to construct mental images from verbal descriptions, but . . . these images have properties which are similar to those images derived from [visual] perception“ (Denis, Goncalves, Memmi 1995, sinngemäßer Klammerzusatz von N.M.).

Bildhafte und sprachliche Informationen werden zwar in verschiedenen Subsystemen des Gehirns verarbeitet, beide Verarbeitungsprozesse scheinen aber nicht gänzlich unabhängig voneinander zu sein. Im Arbeitsgedächtnis (dem 'Flaschenhals', den jeder Sinneseindruck passieren muß) werden Informationen vielmehr in geometrische (metrische und topologische) und nicht-geometrische Komponenten zerlegt, so daß sie später im Langzeitspeicher mehrfach kodiert vorgehalten werden können. Die geometrische Kodierung dient dazu, sich Informationen schneller ins Bewußtsein zurückzurufen (sich assoziativ zu erinnern) und stellt damit eine wichtige Lernhilfe auch für nicht-geometrische Informationen dar. Wiederum muß die geometrische Kodierung nicht in Form eines physischen Bildes o.ä. vorgegeben sein, sondern kann ebenso auf einer mentalen Vorstellung beruhen.

Hierin liegt eine sowohl einfache als auch experimentell nachvollziehbare Erklärung für die Schwierigkeiten, die vielfach bei der Vermittlung abstrakter (häufig mathematischer) Sachverhalte auftreten. Beispielsweise leidet jeder Versuch, das methodische Verfahren einer Faktorenanalyse zu erklären (vgl.

Bahrenberg, Giese, Nipper 1992, S: 198ff.), ganz offensichtlich darunter, daß sich die meisten Menschen von einem  $n$ -dimensionalen Raum *kein Bild machen können*: Man kann einen solchen Raum nicht ohne weiteres auf einem zweidimensionalen Blatt Papier veranschaulichen, und entsprechend gelingt auch die mentale Operation (Imagination!) in der Regel nur schlecht (mit der Folge, daß viele Menschen zwar diffus etwas behalten, aber nicht korrekt – da nicht geometrisch-assoziativ – aus ihrem Langzeitspeicher abrufen können).

Die Annahme, daß funktionale Zusammenhänge zwischen räumlichen und nicht-räumlichen Kodierungen existieren, wird auch von Emmorey, Kossly und Bellugi (1993) erhärtet. Sie beobachteten, daß sowohl taube als auch hörende Individuen, die die amerikanische Handzeichensprache ASL (American Sign Language) beherrschen, wesentlich schneller bei der Durchführung mentaler Rotationsaufgaben sind als solche, die ASL nicht beherrschen. Die visio-spatialen Fähigkeiten innerhalb einer funktionalen Domäne (Sprache) können offensichtlich visio-spatiale Fähigkeiten innerhalb einer anderen funktionalen Domäne (Vorstellung) beeinflussen.

Nun drängt sich natürlich die Frage auf, ob mentale Bilder auch die Grundlage für jene ‘gesamträumlichen Repräsentationen’ bilden, die als kognitive Karten bezeichnet werden. Vieles spricht dafür, z.B. die Beobachtung, daß sowohl mentale Karten als auch mentale Bilder sowohl über visuelle als auch über haptische oder narrative Erfahrungen konstruiert werden können. Desweiteren werden Informationen über mentale Bilder doppelt kodiert, ganz ähnlich zur kognitiv-kartographischen Annahme der Kodierung von erdräumlichen und attributiven Informationen. In diesem Sinne ist der (erdräumliche) spatial priming-Effekt nicht mehr als eine geographische dual coding-Variante (statt beliebiger werden erdräumliche Informationen doppelt kodiert).

Eigenartigerweise wird diese Frage (nach Kenntnis des Verfassers) in der geographischen Literatur nicht behandelt, obwohl sich viele Geographen bei der Diskussion kognitiver Karten auf die imagery debate beziehen. Der Grund dafür liegt sicherlich nicht darin, daß der Zusammenhang als so trivial angesehen wird, daß er keines weiteren Kommentars bedarf. Vielmehr ist die geographische Forschung auch auf diesem Gebiet dem Erdraum verhaftet und übersieht, daß sich die imagery debate keinesfalls auf erdraumbezogene (bzw. geographisch relevante) Geometrien konzentriert, sondern ganz im Gegenteil den räumlichen Bezugssystemen überhaupt keine inhaltliche Bedeutung zumißt. Aus kognitionspsychologischer Sicht ist einzig die Beziehung zwischen geometrischen und nicht-geometrischen Informationen und deren mentaler Repräsentation maßgeblich; daß vereinzelt auch ‘real worlds’ untersucht werden, ist ebenso selbstverständlich wie bedeutungslos. Geographisch gesehen ist dagegen ausschließlich die mentale Repräsentation erdräumlicher Informationen von Interesse (vgl. beispielhaft Nelson 1996; Kitchin 1997a).

Abgesehen davon, welchen Erkenntnisgewinn sich Geographen von dieser Art Fragestellung versprechen, ist es nach Ansicht des Verfassers wesentlich fruchtbarer, statt der mentalen Repräsentationen geographisch-räumlicher Informationen die geometrisch-räumlichen mentalen Repräsentationen beliebiger Informationen zu untersuchen. Der Unterschied, der hier zur geographischen Auffassung besteht, liegt nicht in einer völligen Negierung des Konstrukts kognitiver Karten. Es wird jedoch sowohl die Bedeutsamkeit der geographischen Differenz zwischen erdräumlichen und anderen geometrischen Kodierungen als auch die Bedeutsamkeit der (wie auch immer gearteten) Kodierung erdräumlich verorteter Phänomene im Gegensatz zu nicht erdräumlich verorteten Phänomenen hinterfragt. Im Gegensatz zur Geographie, die nach der mentalen Kodierung der ‘Welt’ in Form von kognitiven Karten fragt, um zum einen die Vermittlung geographischen Wissens voranzutreiben (in Form von ‘besseren’ Karten, Lehrplänen oder GIS-Benutzeroberflächen) und zum anderen über die gefundenen Kodierungsschemata auf die ‘Welt’ zurückschließen will, um diese zu verstehen, wird hier also davon ausgegangen, daß die ‘Welt’ in all ihrer ‘Räumlichkeit’ uns zu menschlichen Wesen hat werden lassen, die nicht nur ‘räumlich’ handeln (wie sonst?) sondern auch ‘räumlich’ denken, schlußfolgern und verstehen. Wir



werden später sehen, daß erdräumlichen Kodierungen dabei sehr wohl eine grundlegende Bedeutung zukommen mag (vgl. Kap. 2.3.3). Als offene Fragen verbleiben jedoch zunächst, wie wir uns mit Hilfe mentaler Vorstellungen Bedeutungen erschließen können, ob diese auf räumlichen bzw. bildhaften Aspekten der Vorstellungen beruhen und – wenn ja – warum das so ist.

## 2.3 Deduktion

Befassen wir uns nun mit jenen mentalen Leistungen, die uns die Bedeutung von Informationen erschließen und insbesondere ermöglichen, neue, d.h. nicht direkt über unsere Sinnesorgane perzipierte Informationen aus alten abzuleiten. Wenn wir eine Schlußfolgerung treffen, gehen wir in der Regel von Prämissen aus, d.h. von gegebenen und als gültig betrachteten Informationen. Wenn es gelingt, durch Vereinfachung und Neukombination dieser Informationen eine neue, allein durch die Prämissen nicht gegebene Information abzuleiten, die weder zu den Prämissen, noch zu sich selbst im Widerspruch steht, dann bezeichnen wir diese neue Information als eine gültige (richtige) Schlußfolgerung bzw. Konklusion und das gesamte Verfahren als *Deduktion*. In Falle von semantischen Informationen sprechen wir dann auch von Erkenntnisgewinn (vgl. Johnson-Laird, Byrne 1991, S. 22).

### 2.3.1 Beweistheoretische Erklärungen

Umstritten ist nach wie vor die wissenschaftshistorisch betrachtet bereits recht alte Frage, inwieweit Menschen logisch denken bzw. logischen Regeln folgen, wenn Sie eine Schlußfolgerung treffen, inwieweit es also formale Regeln gibt, nach denen Individuen gegebene Informationen vereinfachen oder neu kombinieren. Die Ansicht, Deduktion basiere auf der Anwendung einer endlichen Menge universaler Prinzipien, die von jedem Individuum eindeutig und fehlerfrei anwendbar seien, läßt sich u.a. zurückverfolgen auf Kants 'Logik' und seinen zwölf reinen Verstandesbegriffen, Booles 'Laws of Thought' oder Leibniz' 'characteristica universalis' und seiner Suche nach dem verbindlichen Algorithmus der Universalwissenschaften. Da logische Fehler dieser Auffassung gemäß prinzipiell unmöglich sind, sind falsche Schlußfolgerungen aufgrund gegebener Prämissen danach entweder richtige Schlußfolgerungen aufgrund von anderen bzw. falschen Prämissen oder aber auf Fehlfunktionen des informationsverarbeitenden Systems zurückzuführen.

Innerhalb der letzten Jahrzehnte war es Piaget, der die Rolle der Logik als Grundlage für menschliche Deduktionsleistungen erneut betont hat. Er argumentiert im wesentlichen, daß Kinder während der körperlichen und geistigen Entwicklung ihr eigenes Handeln internalisieren und darüber reflektieren, und daß sich dadurch (etwa in der Adoleszenz) ein Set der zum Deduzieren notwendigen formalen Operationen herausbildet. Über die Funktionsweise dieser Operationen und die so erworbenen deduktiven Fähigkeiten besteht bei Inhelder und Piaget unzweifelhaft Einigkeit: „No further operations need be introduced since these operations correspond to the calculus inherent to the algebra of propositional logic. In short, reasoning is nothing more than the propositional calculus itself“ (1958, S. 305; vgl. auch Piaget 1954; Henle 1962; Osherson 1975; Braine 1978; Rips 1983; ein konkretes Gegenbeispiel geben Byrne, Johnson-Laird 1989, S. 567-568).

Mit einer solchen *mechanistischen, objektivistischen* oder auch *beweistheoretischen*, unter anderem auch von der frühen KI-Forschung adaptierten Theorie der propositionalen Repräsentation von Wissen und des logischen Schließens wird also versucht, Deduktion als ein syntaxbasiertes algorithmisches Verfahren zu begreifen und eine formale Kodifizierung deduktiver Denkmuster zu entwickeln. Denken ist nach dieser Auffassung vollständig mit Hilfe des prädikatenlogischen Kalküls beschreibbar, und ein denkendes bzw. schlußfolgerndes informationsverarbeitendes System (sei es nun Mensch, Tier oder Maschine) ist eindeutig durch die Symbole und (prädikatenlogischen) Regeln determiniert,

mit denen es operieren kann. Diese Symbole sind exakt und eindeutig, und ihre Transformationsregeln und deren Alphabet können wiederum in exakten und eindeutigen Begriffen erklärt werden. Die Symbole selbst sind frei von Bedeutung und unterliegen als solche ebenfalls inhaltsleeren Transformationsregeln, so daß semantische Phänomene wie ‘Verständnis’ oder ‘Bedeutung’ ausschließlich über die Beziehung der mentalen Symbole zu ‘Dingen’ in der nicht-mentalen ‘realen Welt’ erklärbar sind. Eine objektivistische Theorie des Denkens setzt damit voraus, die Struktur der ‘wahren Welt’ existiere unabhängig von den sie wahrnehmenden Lebewesen und ‘Bedeutung’ entstehe einzig durch die Referenz mentaler Symbole auf das metaphysische Konstrukt einer externen Realität jenseits der menschlichen Erfahrungen (oder anders formuliert:  $Bedeutungen \subseteq mentale\ Symbole \times reale\ Objekte$ ). Die Transformationsregeln gestatten dem informationsverarbeitenden System insbesondere nicht, bei einem beliebigen Schritt von einem ‘Denkzustand’ zum nächsten aus mehreren Alternativen auszuwählen, sondern schreiben präzise vor, wie sich das System bei jeder Zustandsänderung zu verhalten hat. In diesem Sinne wird Denken, analog zu einem mathematischen Beweis, als ein *effektives Verfahren* beschrieben (und der Mensch damit in letzter Konsequenz als eine spezielle Turingmaschine; vgl. z.B. Turing 1937; Lewis, Papadimitriou 1981, S. 168ff.; auch Weizenbaum 1976, S. 65ff.).

### 2.3.2 Mentale Modelle

Als Alternative entwickelte Johnson-Laird (1975) Mitte der siebziger Jahre die Theorie mentaler Modelle und beschreibt damit Denkprozesse als ein im wesentlichen *nicht effektives Verfahren* (vgl. zusammenfassend Johnson-Laird 1980 sowie ausführlich Johnson-Laird 1983, Johnson-Laird, Byrne 1991; auch Garnham 1997; Knauff et al. 1998). Ansatzpunkt dieser Theorie ist die Beobachtung, daß eine objektivistische Betrachtungsweise zwar gültige von ungültigen Konklusionen zu unterscheiden vermag, aber nicht erklären kann, wann und warum ein Individuum eine bestimmte Konklusion (aus der Menge aller gültigen Konklusionen) trifft. Dies soll im folgenden näher erörtert werden.

#### 2.3.2.1 Syllogistisches Schließen und der figurale Effekt

In den meisten empirischen Untersuchungen über den einer Deduktion zugrundeliegenden Mechanismus – und so auch bei Johnson-Laird – werden (in der Tradition Aristoteles’) der Einfachheit halber sog. *three-term series* Probleme behandelt. Dabei handelt es sich um Syllogismen, deren Prämissen innerhalb eines Experiments vorgegeben werden, um die Konklusion den Probanden dieses Experiments zu überlassen. Eine typische Aufgabenstellung besteht z.B. aus der Vorgabe der Prämissen

Einige Eltern sind Kutscher.  
Alle Kutscher sind Geographen.

aufgrund derer ein Proband z.B. die gültige Konklusion

Einige Eltern sind Geographen. (a)

ableiten wird. Diese Konklusion ist natürlich nicht zwingend zu treffen, möglich wäre auch die ebenso gültige Konverse

Einige Geographen sind Eltern. (b)

und darüber hinaus auch Konjunktionen, wie z.B.

Einige Eltern sind Kutscher und alle Kutscher sind Geographen. (c)

oder

Einige Eltern sind Kutscher und alle Kutscher sind Geographen  
und einige Eltern sind Kutscher. (d)

Die letzten beiden Konklusionen (c) und (d) entsprechen natürlich, ungeachtet ihrer logischen Validität, nicht unserer Vorstellung von *Schlußfolgerungen*, da sie keine *neuen* Sachverhalte ausdrücken, sondern die Lage nur komplizieren. Die ersten beiden Konklusionen (a) und (b) drücken dagegen einfache Sachverhalte aus, die nicht explizit in den Prämissen genannt wurden, und sind damit deduktiv im oben genannten Sinne. Allein bei dem Versuch, nachzuvollziehen, warum Menschen (mit Ausnahme der Logiker) fast ausschließlich Konklusionen der Varianten (a) und (b) treffen, geraten Verfechter objektivistischer Theorien in ernsthaften Erklärungsnotstand (vgl. Johnson-Laird 1980, S. 76-79, Johnson-Laird, Byrne 1991, S. 111-112). Experimentelle Beobachtungen zeigen zudem, daß beim o.g. Beispiel etwa 80% aller berichteten und gültigen Konklusionen solche der Variante (a) sind. Diese systematische Abweichung begründet Johnson-Laird mit dem sog. *figuralen Effekt* (Johnson-Laird 1980, S.74, Johnson-Laird, Byrne 1991, S. 112ff.; vgl. auch Rauh, Schlieder, Knauff 1997). Die Begriffswahl wird schnell deutlich, wenn die Syllogismen in kürzerer Form notiert werden. Das obige Beispiel könnte man z.B. mit  $E = \text{Eltern}$ ,  $K = \text{Kutscher}$  und  $G = \text{Geographen}$  schreiben als

$$\begin{array}{ll} E \cdot K & E \cdot K \\ K \cdot G & \text{bzw.} \quad K \cdot G \\ E \cdot G & G \cdot E \end{array}$$

oder allgemeiner als

$$\begin{array}{ll} A \cdot B & A \cdot B \\ B \cdot C & \text{bzw.} \quad B \cdot C \\ A \cdot C & C \cdot A \end{array}$$

Mit dem Begriff des *figuralen Effekts* wird nun nichts anderes beschrieben, als daß Prämissen der Form  $A \cdot B$ ,  $B \cdot C$  in der Regel zu einer Konklusion der Form  $A \cdot C$  und eben nicht der Form  $C \cdot A$  führen, daß also die Konklusion den Prämissen 'figürlich' entspricht. Neben anderen Einwänden liegt hier eines der Hauptargumente Johnson-Lairds. Der figurale Effekt läßt sich nämlich unter objektivistischen Annahmen in keiner Weise mehr erklären, denn „Logical principles can determine whether a given conclusion is valid, but they cannot even in principle specify, what particular conclusion to draw from some premises on a given occasion, because there are always infinitely many valid conclusions that could be drawn“ (Johnson-Laird 1980, S. 80). Mit anderen Worten: Der figurale Effekt dürfte im Falle einer Deduktion nach rein logischen Regeln nicht systematisch sein, und Deduktion muß anders 'funktionieren' als einzig über die Anwendung eines logischen Kalküls. Tatsächliche, individuell deduzierte Schlüsse unterliegen demnach offensichtlich neben dem prädikatenlogischen Regelwerk noch anderen Beschränkungen, die dem jeweils deduzierenden Individuum inhärent sind: „Hence, in order to drive a *specific* conclusion from the premises, you need some extra logical principle to guide you“ (ebd.; Hervorhebung von N.M.).

### 2.3.2.2 Das Drei-Phasen-Modell des Schlußfolgerns

Was genau neben der Anwendung logisch-formaler Regeln für deduktive Fähigkeiten von Bedeutung ist, veranschaulicht Johnson-Laird anhand von Alltagssituationen. Betrachten wir folgende Situation beim Poolbillardspiel (Johnson-Laird, Byrne 1991, S. 35; vgl. auch Johnson-Laird 1980, S. 79ff.): Die Prämissen

- Die schwarze Kugel liegt direkt hinter dem Queueball.
- Die grüne Kugel liegt rechts vom Queueball.
- Zwischen der grünen Kugel und dem Queueball liegt eine rote Kugel.

erlauben einem Billardspieler z.B. folgenden Schluß

- Wenn ich mich so vor den Billardtisch stelle, daß die rote Kugel zwischen mir und der schwarzen Kugel liegt, dann liegt der Queueball von mir aus gesehen links.

Es ist möglich, diesen Schluß aufgrund formaler Regeln abzuleiten (vgl. Johnson-Laird 1975), aber es ist unwahrscheinlich, daß dies wirklich geschieht (der Leser möge vor dem Weiterlesen zunächst versuchen, die Konklusion zu überprüfen). Die textliche Beschreibung der Lage der Billardkugeln in den o.g. Prämissen reicht in der Regel nicht aus, einen gültigen Schluß zu ziehen. Im Gegenteil: Die Fähigkeit, von den Prämissen deduzieren zu können, setzt offensichtlich das Verständnis der beschriebenen Situation voraus. Um dieses Verständnis zu entwickeln, mag es für die einen notwendig sein, die Lage der einzelnen Billardkugeln zunächst auf einem Blatt Papier zu skizzieren, andere können sich die Situation ohne Papier und Bleistift vorstellen. In jedem Fall werden während des Verständnisprozesses die textlichen Beschreibungen der Prämissen – aber ebenso auch sinnliche Wahrnehmungen von Prämissen (vgl. Marr 1982) – sukzessive in eine mentale Vorstellung dessen, was in den Prämissen beschrieben wird (und nicht des Textes/der Wahrnehmung selbst!) transformiert, und sei es über den Umweg einer Skizze: Die Situation ist genau dann *verstanden*, wenn sie *vorstellbar* ist. Die Situation in unserer Vorstellung ist dann natürlich nicht die Situation selbst, sondern eine mentale, abstrakte und/oder bildhafte Beschreibung ihrer wesentlichen Eigenschaften, mit anderen Worten ein mentales Modell der Situation im Sinne Johnson-Lairds. Die Theorie mentaler Modelle besagt nun, daß man, um Schlußfolgerungen z.B. bzgl. der Billardkugeln treffen zu können, zunächst ein eben solches Modell (eine Vorstellung) der Ausgangssituation entwickeln muß, bevor man dann auf der Grundlage dieses Modells über die Situation hinausgehend rasonieren kann: „. . . it seems likely that people will make it by *imagining* the layout of the balls. This idea lies at the heart of the theory of mental models“ (Johnson-Laird, Byrne 1991, S. 35; Hervorhebung von N.M.).

Dieser Schritt der Modellbildung wird in der Terminologie Johnson-Lairds als *comprehension phase*, in der deutschsprachigen Literatur auch als Phase der *Modellkonstruktion* (zuletzt z.B. Rauh et al. 1998, S. 2) bezeichnet. Die Idee, daß ein Organismus ein internes Modell seiner (wie auch immer perzipierten) physischen Umwelt konstruiert, ist nicht neu. So schreibt Craik bereits 1943: „If the organism carries a ‘small-scale model’ of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude with the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more complete manner to the emergencies which face it“ (zitiert nach Johnson-Laird 1980, S. 73). Craik hat mit den Verweisen auf ein ‘Modell in unseren Köpfen’ sowie auf dessen Potential zur Prüfung und Variation von (Handlungs-)Alternativen bereits zwei zentrale Elemente der Theorie mentaler Modelle vorweggenommen. Folgen wir der Theorie nämlich weiter, wird erst nachdem ein mentales Modell existiert in einem zweiten Schritt (der *description phase* bzw. Phase der *Modellinspektion*) versucht, dieses Modell je nach Aufgabenstellung zu vereinfachen und (gewissermaßen durch eine Neuformulierung des Modells) daraufhin zu inspizieren, ob es informative, d.h. neue, in den Prämissen nicht explizit erwähnte Beziehungen enthält. Wird keine solche informative Beziehung (Konklusion) gefunden, dann ist die einzig gültige ‘Schlußfolgerung’ die, daß nichts aus den Prämissen folgt. Gibt es jedoch eine gültige Konklusion, wird in einem dritten Schritt (der *validation phase* bzw. Phase der *Modellvariation*) versucht, solche alternativen Modelle zu bilden, in denen sich diese angenommene Konklusion als falsch herausstellt. Wird kein solches Modell gefunden, wird die Konklusion endgültig als richtig akzeptiert. Andernfalls wird das Verfahren beim zweiten Schritt fortgesetzt, indem untersucht wird, ob es eine Konklusion gibt, die für alle bis dahin entworfenen Modelle Gültigkeit besitzt. Wird eine solche Konklusion gefunden, wird erneut nach Gegenbeispielen gesucht, und das solange, bis die Konklusion entweder nicht falsifiziert werden kann oder alle möglichen Modelle überprüft worden sind.

Zur Veranschaulichung der drei Modellphasen betrachten wir zunächst ein weiteres, diesmal primär nicht-geometrisches Beispiel, und stellen uns vor, wir seien am Theater mit der Besetzung eines neuen Stückes beauftragt, in dem je drei Sänger, Poeten und Professoren auftreten (nach Johnson-

Laird 1980, S.79ff.). Leider haben wir das Stück noch nicht recht verstanden, insbesondere was die Beziehungen zwischen Sängern und Poeten anbelangt. Wir können maximal sechs Schauspieler beschäftigen, und die einzigen Regievorgaben seien

Alle Sänger sind Professoren.

Alle Poeten sind Professoren.

Was folgt nun aus diesen Prämissen für die Beziehungen zwischen den einzelnen Rollen? Zur Klärung könnten wir beispielsweise sämtliche Schauspieler auf der Bühne versammeln und ihnen gemäß der Prämissen ihre Rollen zuweisen. Natürlich werden wir für unseren Klärungsversuch eine vorläufige Heuristik entwickeln, z.B. indem wir möglichst viele Rollen durch möglichst wenig Schauspieler zu besetzen versuchen. Wir bestimmen also drei beliebige Schauspieler, die die Sängerrollen übernehmen und weisen ihnen, der ersten Prämisse folgend, jeweils auch die Rolle eines Professors zu. Die drei anderen Schauspieler sind für den Moment ohne Rolle:

Sänger = Professor

Sänger = Professor

Sänger = Professor

(?)

(?)

(?)

Die zweite Prämisse interpretieren wir ebenso, und gemäß unserem heuristischen Prinzip vergeben wir die Rollen der Poeten in folgender Weise:

Sänger = Professor = Poet

Sänger = Professor = Poet

Sänger = Professor = Poet

(?)

(?)

(?)

Gleich, ob wir diese Konfiguration von Schauspielern und Rollen tatsächlich auf der Bühne (z.B. durch das Verteilen von Pappschildchen) oder nur in unserer Vorstellung (als ‘Gedankenexperiment’) durchgespielt haben, könnten wir jetzt, da wir über ein erstes Modell verfügen (am Ende der ‘comprehension phase’), fälschlicherweise schließen, alle Sänger seien Poeten (‘description phase’; in der Tat wurde dieser Schluß von einigen Teilnehmern dieses Experiments getroffen; vgl. Johnson-Laird, Steedman 1978). Da wir jedoch bei der Premiere keine böse Überraschung erleben wollen, versuchen wir, unseren vorläufigen Schluß zu widerlegen, indem wir nach alternativen Modellen suchen, die zwar nicht zu den Prämissen, aber zum vorläufigen Schluß in Widerspruch stehen (‘validation phase’). Wir werden wahrscheinlich schnell entdecken, daß es ein solches Modell gibt, z.B.:

Sänger = Professor = Poet

Sänger = Professor = Poet

Sänger = Professor

Professor = Poet

(?)

(?)

Aufgrund unseres Alternativmodells ziehen wir jetzt vielleicht den Schluß, einige der Sänger seien Poeten (‘comprehension phase’). Da wir aber wirklich sicher gehen wollen, finden wir ein drittes Modell (‘validation phase’), nämlich

Sänger = Professor  
 Sänger = Professor  
 Sänger = Professor  
           Professor = Poet  
           Professor = Poet  
           Professor = Poet

An diesem Punkt werden wir schließlich wie folgt deduzieren müssen: Wir waren in der Lage, die Rollenverteilung so zu arrangieren, daß alle Sänger auch Poeten waren. Nun haben wir ein Arrangement erreicht, in dem kein Sänger ein Poet ist. Beide Arrangements widersprechen in keiner Weise den Prämissen. Daher gibt es überhaupt keine gültige Schlußfolgerung bzgl. der Beziehungen zwischen Sängern und Poeten. Dieser Schluß kann wiederum als Syllogismus formuliert werden:

Alle Sänger sind Poeten.  
 Kein Sänger ist Poet.  
 ?

Entsprechend der Theorie wird das Rasonieren über diesen Syllogismus bereits in der zweiten Phase mit der Aussage ‘Es gibt keine gültige Konklusion’ beendet werden, da wir nicht in der Lage sind, aus den Prämissen ein widerspruchsfreies Modell zu entwickeln.

### 2.3.2.3 Evidenz mentaler Modelle

Wie kann nun der figurale Effekt mit Hilfe der Theorie mentaler Modelle erklärt werden? Die Antwort liegt in der Art der Konstruktion eines mentalen Modells (‘comprehension phase’). Wie wir oben bereits anhand des Billard-Beispiels gesehen haben, wird ein Modell schrittweise aufgebaut. Genauer: In der Reihenfolge, in der die Prämissen vorliegen bzw. perzipiert werden, wird das Modell um den Inhalt der einzelnen Propositionen, die in den Prämissen genannt sind, erweitert. Liegen Prämissen der Form  $A \cdot B$  und  $B \cdot C$  vor, so wird die zweite Prämisse ( $B \cdot C$ ) der ersten ( $A \cdot B$ ) gewissermaßen über die Proposition  $B$  angehängt, so daß die Propositionen im resultierenden mentalen Modell eine eindeutige Ordnung aufweisen, nämlich  $A \cdot B \cdot C$ . Dem Schluß bzgl. der Propositionen  $A$  und  $C$  liegt diese Ordnung zugrunde, so daß eine Konklusion der Form  $A \cdot C$  und nicht der Form  $C \cdot A$  naheliegt (wobei natürlich implizit vorausgesetzt wird, daß die Reihenfolge, in der die einzelnen Propositionen perzipiert werden, der Reihenfolge entspricht, in der sie – auch längerfristig – zwischengespeichert und wieder erinnert werden; es gibt allerdings keinen Hinweis, das dies nicht so ist).

Ehrlich und Johnson-Laird (1982) untersuchten das dem figuralen Effekt verwandte Phänomen der referentiellen Kontinuität. Prämissen der Form  $A \cdot B$ ,  $B \cdot C$  werden aufgrund der Beziehung zwischen der ersten Prämisse (die mit der Proposition  $B$  endet) und der zweiten (die mit der Proposition  $B$  beginnt) als kontinuierliche Beschreibungen angesehen, so wie in

$A \cdot B$             Der Zucker steht links vom Senf.  
 $B \cdot C$     z.B.    Der Senf steht vor dem Messer.  
 $C \cdot D$             Das Messer liegt links vom Löffel.

Daneben existieren nicht-kontinuierliche Beschreibungen, im einzelnen sowohl semikontinuierliche Beschreibungen, wie in

$A \cdot B$             Das Bier steht hinter dem Wasser.  
 $B \cdot C$     z.B.    Das Wasser steht links vom Brandy.  
 $A \cdot D$             Das Bier steht links vom Scotch.

als auch diskontinuierliche Beschreibungen, z.B.

$A \cdot B$             Die Katze steht vor dem Pferd.  
 $C \cdot D$     z.B.    Der Fuchs steht vor dem Hund.  
 $B \cdot D$             Das Pferd steht rechts vom Hund.

Ehrlich und Johnson-Laird konnten zeigen, daß Sentenzen, in denen die Kontinuität der Objektnennungen unterbrochen ist (die also nicht kontinuierlich sind in dem Sinne, daß auf Objekte verwiesen wird, ohne einen Bezug zu bereits genannten Objekten herzustellen), sehr viel schwieriger vorstellbar und zu erinnern sind als kontinuierliche Beschreibungen. Offensichtlich hat der Grad der referentiellen Kontinuität einen positiven Einfluß auf die Fähigkeit, das Beschriebene zu lernen – ein Befund, der mit einer formal-logischen Theorie nicht erklärbar ist. Die Theorie mentaler Modelle gibt dagegen die einfache Begründung, daß nicht-kontinuierliche Beschreibungen die – der Beschreibung folgende – sequentielle Integration der Informationen zu einem Modell nicht erlauben. Vielmehr müssen in diesen Fällen mehrere Modelle konstruiert und anschließend erneut integriert werden – oder aber, es wird überhaupt kein Modell konstruiert. Das kostet nicht nur Zeit, sondern beansprucht das Arbeitsgedächtnis auch in ungleich höherer Intensität (vgl. auch Mani, Johnson-Laird 1982).

Nach der Theorie mentaler Modelle sind also richtige Schlußfolgerungen um so einfacher zu treffen, je weniger Modelle aufgrund der Prämissen konstruiert werden können. Eine objektivistische Theorie, die Schlußfolgerungen als Anwendung formaler Ableitungsregeln begreift, prognostiziert dagegen, daß Schlußfolgerungsaufgaben dann einfacher zu lösen sind, wenn hierzu weniger Ableitungsschritte benötigt werden. Diese Thesen wurden von Byrne und Johnson-Laird (1989) überprüft. Sie entwarfen u.a. drei Sets von raumbezogenen Prämissen. Die erste Sentenz (Aufgabentyp I)

$A$  ist rechts von  $B$ .  
 $C$  ist links von  $B$ .  
 $D$  ist vor  $C$ .  
 $E$  ist vor  $B$ .  
 Wie ist die Beziehung zwischen  $D$  und  $E$ ?

erlaubt die Konstruktion genau eines Modells

$C$   $B$   $A$   
 $D$   $E$

das zu folgendem richtigen Schluß führt:

$D$  ist links von  $E$ .

Dieses Set von Prämissen kann daher klassifiziert werden als *Ein-Modell-Problem mit eindeutiger Lösung* (man beachte, daß die erste Prämisse für die Schlußfolgerung irrelevant ist). Die zweite Sentenz (Aufgabentyp II)

$B$  ist rechts von  $A$ .  
 $C$  ist links von  $B$ .  
 $D$  ist vor  $C$ .  
 $E$  ist vor  $B$ .  
 Wie ist die Beziehung zwischen  $D$  und  $E$ ?

erlaubt dagegen die Konstruktion von zwei unterschiedlichen Modellen, nämlich

$C$   $A$   $B$                      $A$   $C$   $B$   
 $D$                      $E$                      $D$   $E$

Da in beiden Modellen die Beziehung zwischen  $D$  und  $E$  identisch ist, weitere Modelle mit den Prämissen jedoch nicht vereinbar sind, führen auch sie zu dem richtigen Schluß:

$D$  ist links von  $E$ .

Die Autoren klassifizieren das zweite Set von Prämissen als *Multiple-Modell-Problem mit eindeutiger Lösung* (bei dem die erste Prämisse diesmal von entscheidender Bedeutung ist). In einem dritten Set wurden die Prämissen so gewählt, daß sie keine gültige Schlußfolgerung zulassen. Dieser Aufgabentyp III umfaßt damit *Multiple-Modell-Probleme ohne eindeutige Lösung*. Beide Theorien (sowohl die

beweis- als auch die modelltheoretische) sagen voraus, daß Aufgaben des Typs III am schwierigsten zu lösen sind. Der Schwierigkeitsgrad der Aufgabentypen I und II muß dagegen unter beweistheoretischen Annahmen identisch sein, da Aufgaben beides Typs zu ihrer Lösung – die notwendigen Regeln zur Transformation von Relationen wie ‘links von’, ‘vor’ etc. vorausgesetzt (vgl. beispielhaft Byrne, Johnson-Laird 1989, S. 567-568) – die gleiche Anzahl von Ableitungsschritten erfordern.<sup>4</sup> Unter modelltheoretischen Annahmen wird im Gegenteil vorausgesagt, daß Aufgaben des Typs II, die die Konstruktion mehrerer Modelle erfordern, schwieriger zu lösen sind als solche des Typs I. Byrne und Johnson-Laird konnten zunächst die übereinstimmenden Annahmen bzgl. des Aufgabentyps III experimentell bestätigen. Desweiteren konnten sie beobachten, daß ihre Probanden in allen Varianten des Experiments signifikant mehr richtige Antworten für Aufgaben des Typs I gaben als für solche des Typs II. Die Autoren schließen daher jene Theorien aus, die Deduktion als ein regelbasiertes Verfahren beschreiben.

#### 2.3.2.4 Mysteriöse Logik

Das Johnson-Lairdsche Drei-Phasen-Modell des Schließens ist nicht unumstritten. Einigkeit besteht jedoch größtenteils über den Kernpunkt der Theorie, daß nämlich „Ausschnitte eines realen, sprachlich beschriebenen oder vorgestellten Sachverhalts im Arbeitsgedächtnis mental modelliert werden, d.h. eine integrierte Repräsentation aufgebaut wird. In dieser Repräsentation entsprechen den ‘realen’ Entitäten Symbole auf der mentalen Seite, und ebenso existieren zu den ‘realen’ Relationen zwischen diesen Entitäten mentale Entsprechungen“ (Rauh et al. 1998, S. 1). Gelingt es, eine solche Repräsentation schrittweise aus gegebenen Prämissen zu entwickeln, also als mentales Modell zu integrieren, wird dieses „zu einer vom Gedächtnis als Ganzes behandelten Entität und kann als solches enkodiert später erinnert werden“ (ebd.). Selbst wenn nicht explizit auf die Theorie mentaler Modelle verwiesen wird, werden mentale Repräsentationen doch zumindest grundsätzlich in jenem allgemeinen Sinne gedacht, den Johnson-Laird in seiner ganzen Unschärfe formuliert: „A mental model is an internal representation of a state of affairs in the external world“ (1992, S. 932). Die Allgemeinheit dieser Formulierung ist Programm und verdeutlicht nach Ansicht von Rauh et al. (ebd.) die Intention des Verfassers, die Theorie mentaler Modelle als allgemeines Rahmenmodell des Denkens und (Text-) Verstehens verstanden zu wissen. Nicht ohne Grund wird auch offengelassen, welche Formen die internen Repräsentationen annehmen können, insbesondere, welche physischen (cerebralen und neurologischen) Aspekte beim Denken eine Rolle spielen – eben jene (mysteriösen?) Beziehungen, die zwischen Denken und Gehirn, d.h. zwischen mentalen und cerebralen Prozessen bestehen.

Die beinahe aufreizend allgemeine Definition verweist damit keineswegs auf eine Schwäche des theoretischen Konstrukts mentaler Modelle. Vielmehr wird deutlich gemacht, worüber die Theorie mentaler Modelle keine Aussagen macht, indem ein Mißverständnis im Ansatz vermieden wird, dem Interpretieren eines beweistheoretischen Ansatzes häufig (gewollt oder ungewollt) aufsitzen: „Inference-rule theories almost certainly have the computational power of universal Turing machines, and so there is unlikely to be any empirical phenomenon that cannot in principle be described within their framework“ (Byrne, Johnson-Laird 1989, S. 574). Mit anderen Worten: Die phänomenal-deskriptive Allmacht regelbasierter Deduktionstheorien folgt direkt aus deren beweistheoretischen Postulaten. Doch mehr noch: Wenn jeder Denkprozeß prinzipiell implementierbar ist (und unter beweistheoretischen Annahmen ist dies in Form einer speziellen Turingmaschine immer möglich: als Computerprogramm, als neuronales Netz, als Gehirn u.v.m.), lassen sich Denkprozesse vollständig auf eine syntaktische – und damit implizit auch auf eine physische – Ebene reduzieren, die, zumindest prinzipiell, verstehbar

<sup>4</sup>Die irrelevante Prämisse in Aufgabentyp I ist eventuell dahingehend irreführend, daß sie zur Ableitung unnötiger Schlüsse bzgl. *A* und *C* genutzt wird; die Ableitung für Aufgaben des Typs I enthielten in diesem Fall zusätzliche Schritte, so daß auch die Voraussage, Aufgabentyp I sei schwieriger als Aufgabentyp II, beweistheoretisch nachvollziehbar ist.



*erscheint*. Die Unterscheidung zwischen einer semantischen Ebene der Deduktion (die Beziehungen zwischen mentalen Symbolen und dem, was sie bedeuten), einer syntaktischen Ebene (den Regeln, nach denen die Symbole transformiert werden können) und einer physischen Ebene (den Mechanismen der Signalverarbeitung) wird damit *scheinbar* obsolet: In gerader Linie kann eine Beziehung zwischen einem Objekt der realen physischen Umwelt und dessen neuronaler Repräsentation in Form synaptischer Ladungspotentiale o.ä. hergestellt werden. Ein beweistheoretisches Konstrukt kommt damit *scheinbar* ohne Mysterien aus. Auf den Kern der hier vorliegenden Täuschung hat bereits von Neumann hingewiesen: „Wenn wir über Mathematik reden, dann diskutieren wir vielleicht eine *Sekundärsprache*, die auf der *Primärsprache* aufgebaut ist, die in Wirklichkeit vom Zentralnervensystem verwendet wird. Somit sind die äußeren Formen *unserer* Mathematik nicht absolut relevant, wenn es um ein Urteil darüber geht, welche mathematische oder logische Sprache vom Zentralnervensystem *wirklich* benutzt wird“ (von Neumann 1958, S. 82; Hervorhebungen im Original). Ähnlich entlarvt Weizenbaum die Funktionsweise dieser Täuschung in moderner Interpretation als „Computermetapher“ (mit der das informationsverarbeitende System ‘Mensch’ durch den Computer versinnbildlicht wird): „Für alle, die sich ganz im Bann der Computermetapher befinden, bedeutet *X* zu verstehen, daß man in der Lage ist, ein Computerprogramm zu schreiben, das *X* realisiert“ (1976, S. 211). Eine mechanistische Theorie setzt damit physische Form mit mentalem Inhalt gleich, indem sie annimmt, das Verständnis der mechanischen Grundlagen sowie der syntaktischen Regeln, mit Hilfe derer Deduktionsprozesse beschrieben werden können, sei das Verständnis des Denkens bzw. Deduzieren selbst.

Folgen wir der Theorie mentaler Modelle, dann treffen wir Schlußfolgerungen, indem wir interne Modelle konstruieren und nach Gegenbeispielen suchen. Die Theorie beschreibt Deduktionsprozesse damit ausschließlich auf semantischer Ebene, beinhaltet keine Aussagen über Schlußregeln und kommt ohne jede formale Logik aus: „The theory contains no rules of inference. Its logical component consists solely in a procedure for testing mental models: the aim is to establish the falsity of a putative conclusion by destroying the model from which it derives . . . The reader will recall that a rule of inference specifies in an essentially ‘syntactic’ way a set of premises and a conclusion that can be derived from them. No such rules are invoked by the theory“ (Johnson-Laird 1980, S. 83). Selbst das Verfahren, nach dem eine Konklusion als ungültig abgelehnt wird, ist also kein formal deduzierter Beweis durch Kontradiktion. Wie oben beispielhaft gezeigt wurde, ergibt sich die Kontradiktion gänzlich ohne Anwendung formaler Regeln gewissermaßen von selbst, indem wir nämlich versuchen, aufgrund aller bislang als gültig erachteten Modelle ein ‘Metamodell’ zu konstruieren. Der Widerspruch folgt dann, wenn wir kein solches Modell finden. Letztendlich wird eine Schlußfolgerung also nicht im formalen Sinne *deduziert*, vielmehr wird sie im Zuge einer sich selbst immer weiter vereinfachenden Modellbildung *gefunden* (oder auch nicht). Interessanterweise beschreibt die Theorie damit Deduktion als rekursiven Prozeß, der eine notwendig induktive Komponente aufweist. Die eigentliche Stärke der Theorie mentaler Modelle liegt somit u.a. darin, daß sie erklärt, warum irren menschlich ist. Oder wie Johnson-Laird und Byrne es beschreiben: „Our view of logical competence is that people are rational in principle, but fallible in practice“ (Johnson-Laird, Byrne 1991, S. 19).

### 2.3.2.5 Mentale Räume

Dem aufmerksamen Leser wird nicht entgangen sein, daß in den Ausführungen zur Theorie mentaler Modelle bei der Auswahl der vorgestellten Beobachtungen und Experimente kaum zwischen raumbezogenen und nicht-raumbezogenen Beispielen differenziert wurde. Es mag der Eindruck entstanden sein, die Entscheidung für oder gegen einen räumlichen Bezug im experimentellen Design fiel gewissermaßen nach Lust und Laune des jeweiligen Experimentators.

Für Untersuchungen, die sich um ein allgemeines Verständnis der strukturellen Eigenschaften mentaler Modelle bemühen, bedeutet die Umstellung von eindimensionalen Syllogismen (‘three-term se-

ries problems') auf komplexere, zweidimensionale Sentenzen ('spatial layouts') zunächst nichts weiter, als daß eine neue Klasse von Relationen eingeführt wird. Neben eindimensionalen Relationen wie 'ist gleich', 'größer als' u.ä. werden nun auch zweidimensionale Relationen betrachtet, z.B. 'links von', 'rechts von', 'oberhalb von' etc. Zweidimensional sind letztere Relationen insofern, als daß sie in sinnvoller Weise konjunkional verwandt werden können, z.B. als 'rechts oberhalb von'. Eines der wichtigsten Kriterien für die Wahl solcher zweidimensionaler experimenteller Problemstellungen ist der höhere Komplexitätsgrad, mit dem Probanden im Experiment konfrontiert werden können. Beispielsweise nehmen Johnson-Laird und Byrne für einen experimentellen Beleg der Evidenz der Theorie mentaler Modelle im Gegensatz zu beweistheoretischen Ansätzen von der Untersuchung einfacher Syllogismen deshalb Abstand, weil „...despite many experimental studies these problems are too impoverished to discriminate among competing accounts of how people solve them“ (Johnson-Laird, Byrne 1991, S. 87). Diesen Punkt führen die Autoren an anderer Stelle weiter aus: „An unfortunate fact about the study of three-term series problems is that it has not yielded any clear answer to the question of which of these two sorts of theory – the proof-theoretic or the model-theoretic – is correct. Moreover, no such answer seems likely to be forthcoming from studying these problems. They are so simple that the two theories appear to mimic one another in their predictions, and so it may be impossible to design a crucial experiment. A more promising domain concerns a richer sort of relational reasoning that has been hitherto neglected: spatial inference“ (Byrne, Johnson-Laird 1989, S. 565).

Eine zweidimensionale (räumliche) Komponente wird von den Autoren also genau deshalb eingeführt, weil erst sie ein für die Überprüfung von Voraussagen konkurrierender Theorien hinreichend komplexes experimentelles Design ermöglicht. Eine solche 'Verräumlichungsstrategie' ist typisch für einen großen Teil der experimentellen Kognitionswissenschaften, die dem Paradigma mentaler Modelle verbunden sind. Andere Autoren konkretisieren dies z.B. folgendermaßen: „Texts that describe spatial layouts can often be constructed so that the propositional structure of the text is different from the spatial structure of the situation described by the text. That is, when objects are close in space, the words used to refer to and describe those objects need not to be close (spatially, temporally, or propositionally) in the description. Thus, demonstrating an effect on comprehension of the structure of the situation indicates that it must be considered in addition to the structure of the description“ (Langston, Kramer, Glenberg 1998, S. 247; vgl. auch Garnham 1992). Forschungsgegenstand ist hier die Differenz zwischen der Struktur der Beschreibung und der Struktur des Beschriebenen. Die Fokussierung auf räumliche Strukturen ist dabei weder inhaltlich durch die in Frage stehenden Thesen bedingt, noch einem primären Interesse an räumlichen Strukturen zuzuschreiben, sondern schlicht ein für die experimentelle Überprüfung (primär nicht-raumbezogener) kognitionswissenschaftlicher Thesen adäquates Mittel und damit von rein forschungstechnischer Relevanz.

Für unsere Zwecke ist natürlich die Frage von Bedeutung, inwiefern mentale Modelle einen 'räumlichen' Charakter haben, inwiefern sie also – unabhängig von der Bildhaftigkeit des modellierten Gegenstandes – bildhaften mentalen Vorstellungen entsprechen. Johnson-Laird (1983, S. 423) faßt mentale Bilder als Teilmenge mentaler Modelle auf, nach Ansicht anderer Autoren sind alle mentalen Modelle im wesentlichen bildhafte Repräsentationen (Guyote, Sternberg 1981), eine Einschätzung, die Johnson-Laird so interpretiert, daß „the ability to construct alternative models should correlate with spatial ability rather than verbal ability“ (Johnson-Laird 1985, S. 190; zit.n. Klauer 1998, S. 12). Interessanterweise wird also zumindest der 'spielerischen' Komponente mentalen Modellierens, nämlich der Suche nach Alternativmodellen während der Phase der Modellinspektion, eine bildhafte Basis zugesprochen. Klauer faßt den Forschungsstand so zusammen: „Insgesamt liegt nahe, daß visuell-räumliche Ressourcen des Arbeitsgedächtnisses beim Konstruieren und Manipulieren mentaler Modelle involviert sind ... [und] daß teilweise bildähnliche mentale Repräsentationen beim Denken konstruiert werden“ (Klauer 1998, S. 12). Neuere Zweitaufgabenexperimente führen Klauer schließ-

lich zu dem Schluß, daß bildhafte Vorstellungen nicht nur bei der Lösung räumlicher Aufgaben eine signifikante Rolle spielen, sondern daß auch in abstrakteren mentalen Modellen, so wie sie für das aussagenlogische Schließen angenommen werden, Relationen zwischen den in Frage stehenden Objekten „durch die räumlich-geometrische Konfigurationen entsprechender Platzhalter“ (Klauer 1998, S. 17) repräsentiert werden.

Interessanterweise wird die Evidenz mentaler Modelle u.a. auch mit einem raumbezogenen Argument begründet, nämlich dem figuralen Effekt. Schon Aristoteles war dieser Effekt bekannt und er bezeichnete einen Syllogismus mit Prämissen der Form  $A \cdot B$ ,  $B \cdot C$  und der Konklusion  $A \cdot C$  als einen perfekten Syllogismus, weil die Transitivität der Beziehung zwischen  $A$ ,  $B$  und  $C$  unmittelbar *sichtbar* sei (Johnson-Laird 1975, S. 75).

### 2.3.3 Kognitive Semantik

Unter der objektivistischen Annahme syntaktischer mentaler Verfahren beruhen mentale Prozesse (und insbesondere Schlußfolgerungen) einzig auf der Manipulation von inhaltsleeren Symbolen, und ihre Bedeutung erlangen sie allein über die Referenz dieser Symbole auf Elemente einer ‘realen Welt’. Dies mag in gewissem Sinne auch auf die Theorie mentaler Modelle zutreffen – je nachdem, wie wörtlich zu nehmen man die Johnson-Lairdsche Unterscheidung von ‘interner Repräsentation’ und ‘externer Realität’ bereit ist. Zwar grenzt sich Johnson-Laird deutlich ab von beweistheoretischen Konstrukten, indem er Denkprozesse explizit als nicht-algorithmische Verfahren deutet. Dennoch operiert seine Theorie mit Symbolen, die in Form von mentalen Modellen immerhin hoch integriert sein können, deren Bedeutung sich aber über ihre Beziehungen zu Objekten einer externen Realität erschließt.

Wie absurd eine objektivistische Annahme – zumindest in ihrer trivialen Interpretation – ist, zeigt z.B. die (Nicht-)Existenz der Goodman’schen Fabelwesen: „... we know a good deal about centaurs and about Don Quixote. And yet, and yet, there are no centaurs, there is no Don Quixote“ (Goodman 1990, S. 359). Ganz offensichtlich ist nicht jedes Objekt, das wir mental modellieren und uns vorstellen, d.h. symbolhaft mental repräsentieren können (von dem wir uns *ein Bild machen können*), notwendig ein Element der Menge aller Objekte einer physisch-realen Welt (und unter beweistheoretischen Annahmen wären wir zudem nicht einmal in der Lage, zwischen jenen ‘irrealen’ Objekten zu unterscheiden: „When we say that centaurs differ from unicorns and from Don Quixote we don’t mean to be taken up our word; for strictly, since none of these exists, all centaurs are unicorns, all unicorns are centaurs, and Don Quixote is both, unicorn and centaur“; ebd., S. 360). Um die Idee mentaler Modelle im Sinne Johnson-Lairds zu retten, bleibt uns nur, die strikte Dichotomie zwischen interner Repräsentation und externer Realität aufzugeben. Dies gelingt leicht, wenn wir eine weitere selbstbezügliche Komponente mentaler Modelle anerkennen: Die externe Realität interner Repräsentationen kann offensichtlich auch durch andere interne Repräsentationen konstituiert werden. Mit anderen Worten: An der *Existenz* sowohl von Menschen als auch Pferden ist nicht zu zweifeln, ebensowenig, wie an der *Vorstellung* eines Zentauren. Aber was liegt dazwischen?

Wir können zunächst festhalten, daß der Mensch sich mit seinem Denken, Vorstellen, Schlußfolgern etc. auf weit mehr zu beziehen vermag als auf eine sinnfreie, externe Realität. Die Entstehung von Vorstellungen und Bedeutungen muß daher auf mehr beruhen als auf Mengenoperationen der o.g. Art. Lakoff (1987, 1988) und Johnson (1987) geben eine mögliche Antwort in Form ihres Konzeptes einer *experientialist cognition*. Ihr Ansatz betont in erster Linie die Rolle des menschlichen Körpers, der körperlichen (Selbst-)Wahrnehmung und der körperlichen Interaktion mit Objekten der physisch-realen Welt als grundlegend für alle relevanten semantischen Phänomene. Ihr Erfahrungsbegriff findet dabei in einer sehr weitgefaßten Bedeutung Verwendung, von sensorischen und motorischen bis hin zu emotionalen und sozialen Erfahrungen, insbesondere aber im Sinne von *angeborenen* Fähigkeiten, die

Wahrnehmungsleistungen aller Art erst ermöglichen. Die Autoren betonen zudem, daß ihr Erfahrungsbegriff nicht in einem empirizistischen Sinne als bloße Verarbeitung von Sinneseindrücken gedeutet werden darf; vielmehr meinen sie mit ‘Erfahrung’ das aktive Teilhaben an natürlichen (physischen) und sozialen Umwelten, d.h. das körperliche Erleben des Seins, das für den Menschen Grundlage und Motivation aller sinnhaften und bedeutungsvollen Erfahrungen und Handlungen ist: „We take common human experience – given our bodies and innate capacities and our way of functioning in a real world – as *motivating* what is meaningful in human thought“ (Lakoff 1988, S. 120; Hervorhebung im Original).

### 2.3.3.1 Kinästhetische Erfahrungen und Bild-Schemata

Im Gegensatz zur Annahme einer sinnfreien Realität, aufgrund derer das Wesen der Bedeutung ein fundamental körperloses ist, gehen also sowohl Lakoff als auch Johnson davon aus, daß die Bedeutung von konkreten (oder abstrakten) Wahrnehmungen (oder Interpretationen) untrennbar mit den individuellen Erfahrungen unseres körperlichen Seins verknüpft sind (vgl. im folgenden zusammenfassend Lakoff 1988, S. 132ff.). Bedeutungszuweisungen werden in diesem Sinne verstanden als im wesentlichen metaphorische Projektionen wohlbekannter Aspekte unserer körperlichen und interaktiven Erfahrungen auf abstrakte Strukturen. Unsere kinästhetischen Erfahrungen geben dabei die bedeutungsrelevanten Kategorien in Form sogenannter *Bild-Schemata* vor, deren metaphorische Projektion sich in einer systematischen Polysemie auf Grundlage des jeweiligen Schemas widerspiegelt. Johnson (1987) führt eine Fülle von Beispielen für derartige Schemata an:

- Das *Container-Schema* bzw. *Innen-Außen-Schema* erlaubt die grundsätzliche Unterscheidung zwischen Innen und Außen. So verstehen wir unseren eigenen Körper sowohl als Container, in den wir z.B. Luft ein- und aus dem wir Luft ausatmen, als auch als Inhalte von Containern, z.B. als Koch innerhalb der Küche. Die metaphorischen Projektionen dieses Schemas führen zu vielfältigen Bedeutungen: Wir gehen *in* das Bett und fallen *in* tiefen Schlaf; wir räumen ein Zimmer oder einen Fehler *ein*; wir schrauben Glühbirnen und geraten selbst *aus* der Fassung; wir holen Leinen und Informationen *ein*; wir führen Hunde und Befehle *aus* usf.
- Das *Teil-Ganzes-Schema* beruht auf den kinästhetischen Erfahrungen, die uns unseren Körper als Ganzes mit manipulierbaren Teilen wahrnehmen lassen: Unser Kopf ist ein *Teil* von uns, wir sind ein *Teil* unserer Familie; wir *teilen* mit Freunden das *ganze* Brot und das *ganze* Leben; die *ganze* Welt ist *aufgeteilt* in Galaxien usf.
- Zentraler Aspekt des *Link-Schemas* ist die metaphorische Projektion von Verbindungen. Unsere erste externe Verbindung ist die Nabelschnur, während unserer frühen Kindheit halten wir uns an unseren Eltern fest, um entweder deren oder unsere (relative) Position zu sichern. Um Dinge relativ zueinander an ihrem Platz zu halten, benutzen wir Schnüre, Seile, Klammern oder andere Mechanismen, die eine physische Verbindung gewährleisten. Die metaphorische Polysemie ist sehr reichhaltig: Wir *halten* unsere Eltern und unsere Versprechen; wir *knüpfen* Teppiche und soziale Kontakte; wir *brechen* Regeln und die *Fesseln* der Gewalt usf.
- Das *Quelle-Weg-Ziel-Schema* entsteht aus der kinästhetischen Erfahrung körperlicher Ortsveränderungen, die jeweils aus einem Ort, von dem wir ausgehen, einem Ort, zu dem wir uns hin bewegen und einem Weg, den wir zurücklegen, bestehen. Auch dieses Schema bietet eine Fülle metaphorischer Mehrdeutigkeiten: Wir *folgen* Wegweisern, Bauanleitungen und Diskussionen; wir *verfolgen* Diebe und politische *Ziele*; wir begeben uns auf Reisen und auf Holzwege; *ausgehend* von Prämissen geben wir Lösungen *näherungsweise* an; die Räder unseres Autos und unsere Beifahrer drehen *durch* usf.

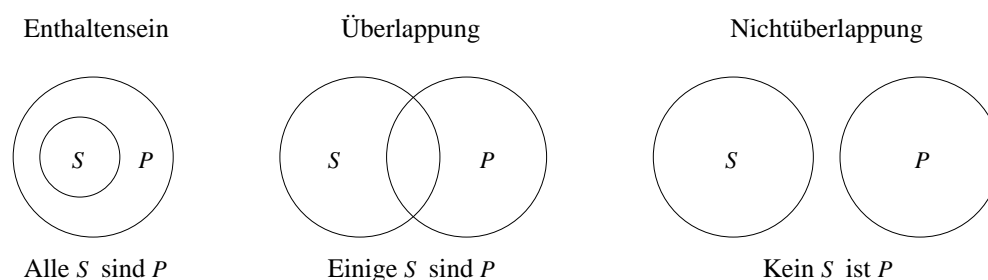
Als weitere bedeutungsrelevante Schemata werden z.B. das *Oben-Unten-Schema* (alpiner und gesellschaftlicher *Auf* - oder *Abstieg*) oder das *Vorne-Hinten-Schema* (*Vor* - und *Nachteil*, *Vorgesetzte* und *Hintermänner*) genannt, zahlreiche andere Schemata sind denkbar. Das zentrale Argument von Lakoff und Johnson ist in jedem Fall, daß abstraktes Denken und Schlußfolgern (1.) auf konkreter körperlicher Erfahrung und (2.) auf deren metaphorischer Projektion in eine abstrakte Domäne beruht. Die Erfahrungsstrukturierung aufgrund von Bild-Schemata ist nicht nur unabhängig von kurzfristigen Wahrnehmungen und vorgefaßten Konzepten, sondern eine angeborene Fähigkeit jedes Einzelnen. Die Art der metaphorischen Projektion ist wiederum keineswegs zufällig, sondern ist selbst motiviert von den unseren täglichen körperlichen Erfahrungen inhärenten Strukturen. Insbesondere werden durch die metaphorische Projektion kinästhetischer Erfahrungen deren grundlegende logische Eigenschaften bewahrt.

### 2.3.3.2 Logische Diagramme

Die letzte Bemerkung verdient kurz eine nähere Erläuterung. Die interne Struktur unserer kinästhetischen Erfahrungen und der daraus resultierenden Bild-Schemata wird von Lakoff (1987) in Form einer *basic logic* beschrieben. Das Container-Schema entspricht z.B. dem Gesetz vom ausgeschlossenen Dritten (etwas ist entweder innerhalb oder außerhalb eines Containers) und folgt dem *modus ponens* (wenn ein Objekt  $x$  innerhalb von Container  $A$  und  $A$  innerhalb von Container  $B$  ist, so ist  $x$  innerhalb von  $B$ ). Desweiteren sind die dem Container-Schema zugrundeliegenden Relationen  $R \in \{\text{'innerhalb von' , 'außerhalb von'}\}$  transitiv: Wenn Container  $A$   $R$  Container  $B$  ist und  $B$   $R$  Container  $C$ , dann ist  $A$   $R$   $C$ . In diesem Sinne transitiv sind auch die anderen genannten Schemata, wenn man die geeigneten Relationen einsetzt (Teil-Ganzes-Schema:  $R = \text{'ist Teil von'}$ ; Link-Schema:  $R = \text{'ist verbunden mit'}$ ; Quelle-Weg-Ziel-Schema:  $R = \text{'von ... führt ein Weg nach'}$ ; Oben-Unten-Schema:  $R \in \{\text{'ist oberhalb von' , 'ist unterhalb von'}\}$ ; Vorne-Hinten-Schema:  $R \in \{\text{'ist vor' , 'ist hinter'}\}$ ). Darüber hinaus ist das Link-Schema symmetrisch (wenn  $A$  verbunden mit  $B$  ist, dann ist auch  $B$  verbunden mit  $A$ ), alle anderen genannten Schemata sind dagegen asymmetrisch. Eine Sonderstellung nimmt das Quelle-Weg-Ziel-Schema ein, das nämlich im nicht-metaphorischen, also räumlich-geometrischen Gebrauch (im Sinne eines ungerichteten Straßengraphen – sehen wir also von 'Einbahnstraßen' und ähnlichen Restriktionen ab) sehr wohl symmetrisch ist.

Nach Lakoff und Johnson erfassen wir z.B. klassische Kategorien und Mengen ebenso wie Aussagen intuitiv als abgeschlossene Räume: Ein Objekt befindet sich entweder innerhalb oder außerhalb einer Kategorie oder wird ihr zugeordnet bzw. aus ihr entfernt; ist  $x$  innerhalb der Kategorie  $A$  und  $A$  innerhalb der Kategorie  $B$ , so ist  $x$  innerhalb der Kategorie  $B$ . Die Logik klassischer Kategorien entspricht also der Logik von Containern (vgl. im folgenden Lakoff 1990; auch May 1995). Nach der hier zugrundeliegenden Container-Metapher ('klassische Kategorien sind Container') werden die logischen Eigenschaften klassischer Kategorien von den logischen Eigenschaften von Containern 'geerbt'. Dies wußte bereits Ende des 18. Jahrhunderts Leonhard Euler, der jeden allgemeinen Begriff als abgeschlossenen Raum ansah, der die einzelnen Objekte enthält, die unter diesen Begriff fallen (vgl. auch May 1995, S. 294f.). Dieser abgeschlossene Raum wurde von ihm durch einen Kreis (einem heute sogenannten *Euler-Kreis*) auf einem Blatt Papier veranschaulicht, so daß er das Verhältnis zweier Propositionen, z.B. dem Subjekt und dem Prädikat einer Aussage, graphisch als Überlappung, Enthaltensein oder Nichtüberlappung einzelner Kreise darstellen konnte (Abb. 2.1).<sup>5</sup> Genau diese Notation wird auch von Lakoff übernommen (Abb. 2.2). Das Diagramm basiert auf dem Innen-Außen-Schema und besitzt damit gewissermaßen eine 'eingebaute Logik', so daß man unmittelbar zu sehen scheint, was aus den Prämissen folgt (May 1995). Auch zahlreiche andere Diagrammtypen sind möglich (vgl.

<sup>5</sup>Für die Frage, inwieweit Euler-Kreise eine syllogistische Logik implementieren, vgl. Stenning, Oberlander 1995.



**Abbildung 2.1** Euler-Kreise. Die drei grundlegenden Konfigurationen von Euler-Kreisen für ein Subjekt  $S$  und ein Prädikat  $P$  (nach May 1995, S. 295).

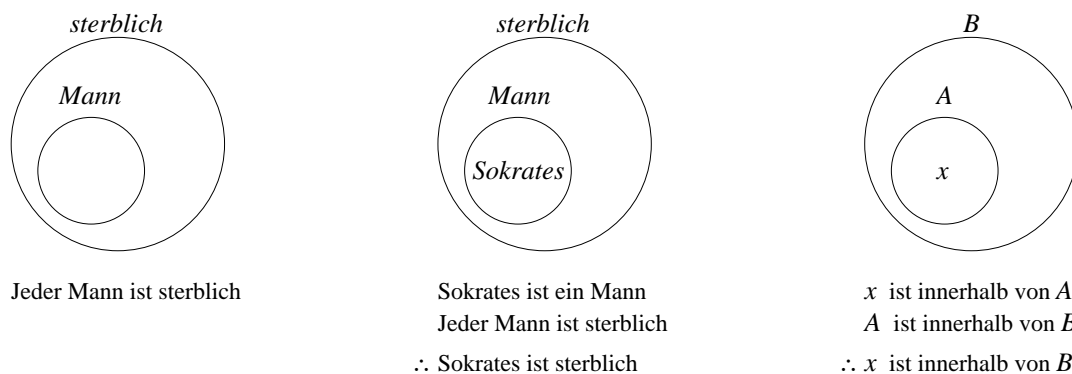
ausführlich Lakoff 1990), deren bekanntester Vertreter sicherlich das Venn-Diagramm ist, das sich als didaktische Hilfe bei der Vermittlung der Mengenlehre seit langem bewährt hat (Venn 1880).

Die Kognitive Semantik, die ja die Behauptung aufstellt „daß jedes Denken – in den Wissenschaften ebenso wie im Alltag – auf intuitiv bedeutungsvollen, aber hochgradig schematischen mentalen Modellen und Vorstellungsschemata beruht, in denen sich die körperliche Erfahrung des kognitiven Subjekts niederschlägt“ (May 1995, S. 286), bietet damit einen weiteren und sehr direkten Zugang zu unserer Frage, inwieweit menschliches Denken auf räumliche Informationen angewiesen ist oder auf ihnen basiert. Offensichtlich liegt in der (sowohl notwendigen als auch trivialen) ‘Räumlichkeit’ der Lebensverhältnisse des Menschen als mobiler Bewohner der Erdoberfläche eine tiefere Bedeutung für seine Denk- und Deduktionsweisen, denn „Mechanisms developed for perceiving and reasoning about the spatial world are likely to be used for reasoning about other domains“ (Stenning, Oberlander 1995, S. 136). Die geographische Erdräumlichkeit scheint demnach eine viel weitergehende Funktion für die menschliche Existenz, sein Verhalten und seine Schlußfolgerungen zu spielen als vielfach (wohl auch von Geographen) angenommen wird.

Der eigentliche Kern der Idee einer kognitiven Semantik liegt also in den Annahmen, daß Menschen (1.) aufgrund ihrer leiblichen Erfahrungen gewissermaßen ‘in Räumen’ denken, daß diese Räume (2.) topologisch identisch sind mit physisch erfahrbaren Räumen und sich daher (3.) auch in ‘reale Räume’ – z.B. in Diagramme, aber auch Karten u.v.m. – zurückübersetzen lassen. Wie gezeigt wurde, finden die ersten beiden Annahmen innerhalb der Kognitionswissenschaften eine breite Bestätigung. Die mentalen Prozesse, die diesem ‘räumlichen Denken’ zugrundeliegen, sind zwar noch vielfach ungeklärt, jedoch erscheint eine sowohl bildhafte als auch nicht-algorithmische Strategie evident. Letztendlich liegt damit der umgekehrte Schluß nahe, daß bildhafte und nicht-algorithmische Vorlagen helfen können, die Konstruktion mentaler Modelle zu erleichtern. Genau das soll im folgenden Kapitel näher untersucht werden.

## 2.4 Diagrammatisches Denken

Unter einem *Diagramm* verstehen wir im folgenden einen analogen (d.h. direkten), zweidimensionalen Repräsentationsmechanismus, der durch eine weitestgehende Übereinstimmung (jedoch nicht unbedingt durch eine Isomorphie) zwischen der Struktur der Repräsentation und der Struktur des Repräsentierten charakterisiert ist. Wir benutzen Diagramme, um Informationen und Daten zu veranschaulichen, und *diagrammatisches Denken* nutzt eine solche visuelle Repräsentation als primäre Quelle für Schlußfolgerungen. Der Begriff des diagrammatischen Denkens geht auf Charles S. Peirce zurück, der bereits zu Beginn dieses Jahrhunderts gezeigt hat, daß Diagramme keine leicht verzichtbaren Illustrationen sind, die wissenschaftliche Texte begleiten, sondern daß sie einen eigenen Typus wissen-



**Abbildung 2.2** Lakoffs Interpretation klassischer Kategorien als Container. Links sind allgemeine Begriffe durch Euler-Kreise dargestellt, in der Mitte zusätzlich Sokrates, der unter diese Begriffe fällt, und rechts die allgemeine Notation (vgl. Lakoff 1990, S. 52).

schaftlichen Verstehens charakterisieren (May 1995). In jüngerer Zeit wurde der Begriff in Arbeiten zur künstlichen Intelligenz und verwandten Bereichen der Computer- und Kognitionswissenschaften adoptiert. Das überrascht insofern nicht, als daß das Rasonieren aufgrund visueller Repräsentationen im allgemeinen als wichtige, aber besonders schwer zu durchschauende und zu simulierende menschliche Fähigkeit gilt (Arnheim 1969; Sloman 1971; Kosslyn 1980; Glasgow, Narayanan, Chandrasekaran 1995; Schlieder 1996, 1997; Anderson 1997; Kulpa 1997). Ein anschauliches Beispiel für ein Diagramm im oben genannten allgemeinen Sinne ist z.B. eine geographische Landkarte, auf der die relative Lage bestimmter Symbole der relativen Lage von Städten auf der Erdoberfläche entspricht. Im Gegensatz dazu ist dies bei einer propositionalen Repräsentation, z.B. einer textlichen Beschreibung der Verteilung von Städten, nicht der Fall (vgl. Kulpa 1994).

Die Informationen, die mit einem Diagramm dargestellt werden, können immer auch in Form einer textlichen oder tabellarischen Beschreibung vermittelt werden. Der Unterschied beim Interpretieren von Diagrammen und propositionalen Repräsentationen liegt daher nicht in der dargebotenen Information, sondern in den Informationsverarbeitungsmöglichkeiten und -regeln (Schnotz 1992). Im allgemeinen werden die Vorteile, den Diagramme gegenüber propositionalen Repräsentationen bieten, wie folgt zusammengefaßt:

- Diagramme beschreiben verwandte Informationen in der Regel in Form einer räumlichen Gruppierung, so daß langwierige Suchoperationen nach verwandten Informationen vermieden werden (Larkin, Simon 1987).
- In Kombination mit anderen Repräsentationsformen (z.B. Text) dienen Diagramme zum einen der Wiederholung von Informationen, zum anderen bieten sie die Möglichkeit zur doppelten Kodierung (im Sinne der dual-coding-Theorie; Paivio 1986; vgl. Glenberg, Langston 1992)
- Diagramme erlauben die direkte Ableitung von Informationen, die auf Basis anderer Darstellungsformen häufig nur implizit gegeben sind. So sind z.B. für die Zeitreihen der Einwohnerzahlen zweier Städte im Liniendiagramm sowohl die Einwohnerzahlen selbst als auch deren Verlauf, deren Differenz sowie die Entwicklung des Verlaufs oder der Differenz unmittelbar sichtbar. Liegen die Zeitreihen dagegen tabellarisch vor, können derartige sekundäre Informationen nur unter großem Aufwand hergeleitet werden (vgl. Larkin, Simon 1987; Schnotz 1992, S. 22ff.; Kulpa 1994).

Die Nutzungseffizienz von Diagrammen ist natürlich keine Eigenschaft der Diagramme selbst, sondern ist Folge einer Interaktion mit dem menschlichen kognitiven Apparat (Schnotz 1992, S. 27ff.). Der Übergang vom einfachen Wahrnehmen eines Diagramms zu dessen Verständnis kann nur erfolgen, wenn dem Diagramm und seinen Teilen spezifische Bedeutungen zugemessen werden. Wenn wir die Vorstellung mentaler Modelle akzeptieren, dann liegt die eigentliche Stärke von Diagrammen in der Möglichkeit, sie als Konstruktionshilfe mentalen Modellierens einzusetzen. In diesem Sinne spricht Fenk (1994) von Diagrammen als Transformationen räumlicher Metaphern in visuelle Analogien, Gyselinck und Tardieu (1994) sehen in ihnen eine mögliche Form des externen Ausdrucks von Teilen mentaler Modelle. Nach Schnotz, Picard und Henninger (1994) dienen Diagramme schließlich zur Strukturierung mentaler Modelle, da dem Verständnis von Diagrammen ein Abbildungsprozess sichtbarer räumlicher Verhältnisse auf semantische Relationen in einem mentalen Modell zugrundeliegt (Schnotz, Picard, Henninger 1994). Ähnlich argumentieren Bauer und Johnson-Laird (1993) oder Scaife und Rogers (1996; vgl. auch Ottoson 1988; Gombrich 1990; Glenberg, Langston 1992; Bryant, Lanca, Tversky 1995; Chen 1995; Do, Gross 1995; DeLoache et al. 1998).

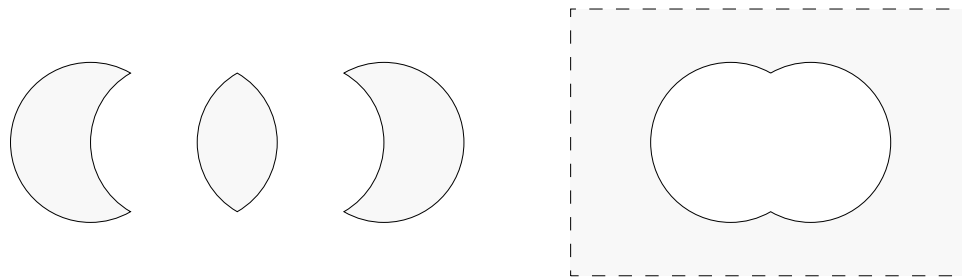
Je ähnlicher die strukturellen Eigenschaften des Diagramms (der Repräsentation) und des Repräsentierten sind, desto weniger wird das Arbeitsgedächtnis mit der Modellkonstruktion belastet und desto einfacher ist auch das mentale Modell selbst strukturiert sowie sein semantischer Gehalt zu erschließen (im Vergleich zu anderen, potentiell konstruierbaren Modellen auf Grundlage derselben, aber anders dargestellten Informationen). Folgen wir der Kognitiven Semantik, dann müssen gerade solche Diagramme mental leicht modellierbar sein, die sich an die kinästhetischen Kategorien der Bild-Schemata von Johnson (1987; vgl. S. 44) anlehnen. Aus dieser Perspektive scheint es zumindest kein Zufall zu sein, daß diagrammatische Umsetzungen des Innen-Außen-Schemas in Form von Euler-Kreisen und Venn-Diagrammen bereits seit Jahrhunderten logische Abhandlungen bebildern.

Das Venn-Diagramm stellt dabei eine notwendige Weiterentwicklung der Euler-Kreise dar, da Euler-Kreise eine gültige und konsistente Interpretation logischer Aussagen nicht in jedem Fall gewährleisten. May (1995) stellt insbesondere die Notwendigkeit heraus, die Interpretationen solcher Diagramme explizit anzugeben. In der gängigen Interpretation sehen wir Euler-Kreise als Umrisse von Teilflächen der Euklidischen Ebene an (vgl. Abb. 2.1 rechts). Der in  $A$  liegende Punkt  $x$  liegt somit sowohl in  $A$  als auch in  $B$ . Interpretieren wir die Abbildung dagegen als zweidimensionale Projektion eines dreidimensionalen Torus, dann bezeichnet  $B$  die Oberfläche des Torus und  $A$  den Bereich, den  $B$  umgibt. Der Punkt  $x$  liegt nach dieser Interpretation in  $A$ , jedoch niemals in  $B$ . Eine Interpretation der Euler-Kreise als Umrisse von Teilflächen der Erdoberfläche (im Sinne der kinästhetischen Erfahrungswelt von Johnson und Lakoff) führt schließlich zu einer dritten Interpretation, die in Anlehnung an May folgendermaßen illustriert werden soll: Stellen wir uns einen Fisch ( $x$ ) vor, der in einem See ( $A$ ) schwimmt, der wiederum in einem Wäldchen ( $B$ ) liegt. Im Gegensatz zu Sokrates, der sowohl Mann und auch Mensch ist, schwimmt unser Fisch zwar im See, jedoch nicht bzw. nur indirekt im Wäldchen (nach May 1995, S. 290/291).

Derartige Mehrdeutigkeiten können weitestgehend vermieden werden, wenn die geometrischen Kombinationsmöglichkeiten der Kreise klar definiert sind. Genau das erreichte Venn (1880) mit einer zweidimensionalen Konfiguration, die aus nur vier geometrisch disjunkten Gebieten besteht (Abb. 2.3). Unter der Annahme zweier Wahrheitswerte lassen sich alle 16 verschiedenen zweistelligen Relationen als eindeutige Kombination dieser vier Flächen darstellen. Im Unterschied zur gängigen Interpretation von Euler-Kreisen wird hier die Euklidische Ebene partitioniert, d.h. die einzelnen Teilgebiete überlappen sich nicht (vgl. auch die Einführung von Kap. 3). Damit wird jeder geometrischen Inklusions- bzw. Exklusionsbeziehung eindeutig je eine der möglichen zweistelligen Relationen zugeordnet.

Wenn wir nun vor diesem Hintergrund die Funktionsweise von GIS betrachten, stellen wir fest, daß die für GIS so elementare Operation der Flächenverschneidung offensichtlich die Projektion des





**Abbildung 2.3** Grundkonfigurationen eines Venn-Diagramms. Das klassische Venn-Diagramm besteht aus vier disjunkten Flächen, aus denen sich – unter der Annahme zweier Wahrheitswerte – die 16 möglichen zwei-stelligen Relationen zusammensetzen lassen. Die Schraffur dient nur zur Markierung der Gebiete, nicht zur Darstellung einer Verneinung; die Darstellung der Teilflächen als getrennte Stücke dient nur der Unterscheidung (nach May 1995, S. 295).

Vennschen Gedankens auf die Erdoberfläche ermöglicht. Aus kognitionswissenschaftlicher Sicht ist diese Vererdräumlichung geradezu ideal, denn sowohl das repräsentierte Objekt (ein Teil der Erdoberfläche) als auch die Struktur der Repräsentation (als Karte) liegen in einem Erfahrungsbereich, den Menschen seit ihrer Frühzeit kulturell tradieren. Dieser Erfahrungsbereich umfaßt sowohl die zur zielgerichteten Mobilität notwendigen Techniken (Orientierung, Navigation, Kartographie) als auch körperlich-dingliche Erfahrungen mit der erdräumlichen Umwelt (auf der Erdoberfläche stehen zwei verschiedene Dinge nebeneinander, manchmal übereinander, aber niemals an genau derselben Stelle). Die mentale Modellierung von erdräumlichen Verschneidungen erfordert daher einen minimalen Abstraktionsaufwand und eröffnet gleichzeitig die Möglichkeit, das Potential der zweiwertigen Logik durch eine ungemein anschauliche Metapher zu erkennen und zu nutzen. Einer der Hauptwesenszüge geographischer Flächenverschneidungen liegt nach Ansicht des Verfassers genau darin, daß sie eine unmittelbar und interkulturell zugängliche Klasse mentaler Modelle eröffnen, die alle, die auf ihrer Grundlage kommunizieren, zur Einhaltung zweiwertiger Logik zwingt.

Eine ähnliche Klasse mentaler Modelle läßt sich schließlich auch erzeugen, wenn die Parallele zwischen Venn-Diagrammen und erdräumlicher Flächenverschneidung für die Verarbeitung solcher Informationen genutzt wird, die nicht im geographischen Sinne erdraumbezogen sind. Die Repräsentation wäre dann keine erdräumliche, sondern eine merkmalsräumliche 'Karte', und die repräsentierten Objekte entsprächen numerischen Daten, die in der Regel nur über formale (nicht-visuelle) Methoden analysiert werden (vgl. Mevenkamp 1996a, 1996b). Gemeint sind also solche Abbildungen, die die aus kognitionswissenschaftlicher Sicht wertvolle diagrammatische Struktur von Flächenverschneidungskarten bzw. Venn-Diagrammen übernehmen und damit eine metaphorische Form erdräumlicher Raumabstraktionen darstellen (vgl. Kap. 4.2).

Überspitzt formuliert liegt die Funktionsweise von geometrischen Abstraktionen, so wie sie mit GIS und dem Verschneideverfahren geboten werden *könnten*, darin, beliebige, flächenhaft darstellbare Informationen mental als ein Venn-Diagramm zu modellieren, das einer erdräumlichen Logik und Metaphorik entspricht. Die Umsetzung dieses Gedankens erfordert die Bestimmung von Partitionierungen auf eine sehr allgemeine Weise (und dies gilt auch für erdräumliche Fragestellungen; vgl. Kap. 4.1). Was hierzu noch fehlt, ist ein geeigneter Verschneidealgorithmus.

## Kapitel 3

# Die Bestimmung planarer Subdivisionen

### 3.1 Problemstellung und Begriffsdefinitionen

Ein Geograph, der mit einer Flurstückkarte ausgestattet die Bodennutzung kartieren will, steht vor der Aufgabe, jedes Flurstück auf seiner Karte mit der tatsächlichen und sichtbaren Nutzung dieses Flurstücks entsprechenden Signatur zu versehen. Wird ein Flurstück als Weide, ein benachbartes als Wiese genutzt, wird dies entsprechend auf der Karte vermerkt. Für den Fall, daß ein Flurstück zum Teil als Weide, zum anderen Teil als Wiese genutzt wird, gibt es für den Geographen zwei Möglichkeiten: Entweder, er teilt das Flurstück auf seiner Karte und vermerkt anschließend die unterschiedlichen Nutzungen auf beiden Teilflächen getrennt, oder er versieht das gesamte Flurstück mit der entsprechenden Signatur der Mischnutzung.

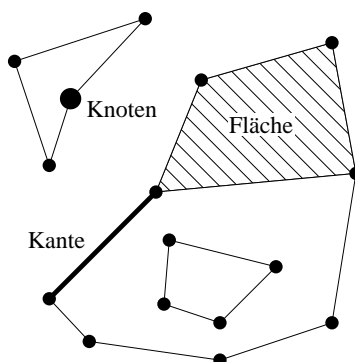
Nach getaner Arbeit wird die Flurstückkarte aussehen wie ein bunter Flickenteppich, der Löcher aufweist z.B. für den Fall, daß Siedlungen und Verkehrswege bei der Kartierung unberücksichtigt geblieben sind. In jedem Fall läßt sich mit einem Blick auf die fertiggestellte Karte jeder kartierten Teilfläche genau eine Nutzung zuordnen; ebenso gibt es keinen Punkt auf der Karte, für den mehr als eine Nutzung vermerkt ist. Sehen wir die nicht kartierten Bereiche – die Löcher ebenso wie die Fläche jenseits des Kartenausschnitts – ebenfalls als Teilflächen mit bestimmter, wenn auch nicht weiter spezifizierter Nutzung an, so können wir schließlich festhalten, daß jeder Punkt der kartierten Fläche zu genau einer Teilfläche gehört, welche wiederum auf genau eine Weise genutzt wird. Es gibt dann weder zwei Teilflächen noch zwei Flächengrenzen, die sich überschneiden oder überlappen. Ebenso wenig gibt es einen beliebigen Punkt, der sich nicht genau einer Teilfläche zuordnen ließe – wenn auch manchmal, im Fall von Löchern, einer ‘unechten’. Mit anderen Worten: Unser Geograph, der sich an der vorhandenen Nutzung orientierte, hat die Erdoberfläche vollständig partitioniert.

In Anlehnung an de Berg et al. (1997, S. 30) bezeichnen wir eine solche Partitionierung der Ebene im folgenden als *planare Subdivision* der Ebene (Abb. 3.1). Eine Subdivision ist planar, wenn der zugrundeliegende Graph planar ist. Dieser Graph besteht aus *Knoten* (Punkten)  $v = (x, y)$ , die durch je ein Koordinatenpaar gegeben sind, und *Kanten* (Linien<sup>1</sup>)  $e = (v_i, v_j)$ , die je zwei Knoten miteinander verbinden. Im Fall von Flächenkarten bezeichnet der zugrundeliegende Graph beispielsweise die Grenzen der betrachteten (Teil-)Flächen. Dieser Graph ist planar, weil er sich ohne Überschneidung seiner Kanten in die Ebene zeichnen läßt. Oder umgekehrt ausgedrückt: Wäre der Graph nicht planar, hätte der Kartograph fehlerhaft gearbeitet, denn er hätte Flächen kartiert, die sich zum Teil überlappen bzw. deren Grenzen sich überschneiden.

Der Graph ist nicht notwendig verbunden, d.h. es können Paare von Knoten existieren, die nicht durch eine Kante oder eine Folge von Kanten (einen Pfad des Graphen) verbunden sind. Eine *Fläche*

---

<sup>1</sup>Kreisbögen u.ä. können durch eine endliche Kanten- bzw. Linienfolge approximiert werden



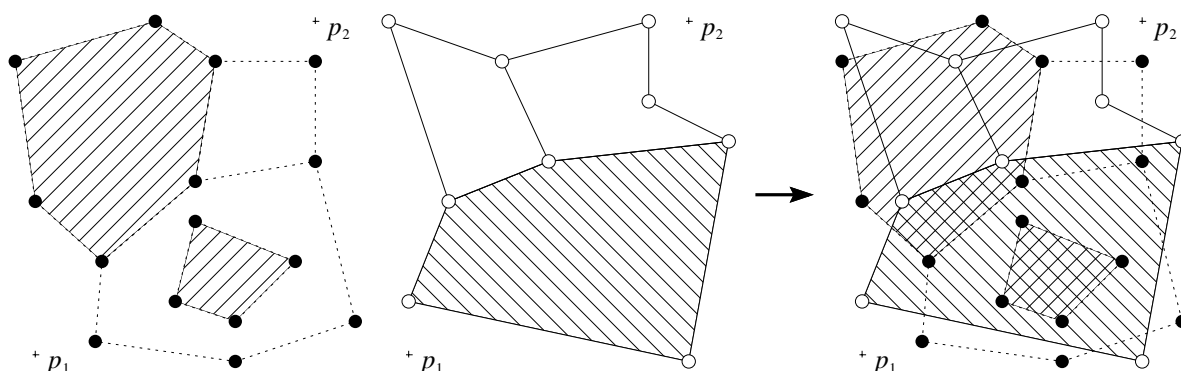
**Abbildung 3.1** Eine nicht verbundene planare Subdivision. Eine vollständige Partitionierung der Ebene wird als planare Subdivision bezeichnet. Die dargestellte Subdivision ist planar, weil es keine zwei Kanten gibt, die sich in einem Punkt überschneiden, der nicht auch ein Knoten ist. Sie ist nicht verbunden, weil nicht jeder Knoten mit jedem anderen Knoten durch eine Kantenfolge verbunden ist.

ist eine Teilmenge der Ebene, die zusammenhängend ist und weder einen Punkt auf einer Kante noch einen Knoten beinhaltet. Eine Teilmenge  $f$  der Ebene ist dann zusammenhängend, wenn jeder Punkt  $p_i \in f$  mit jedem anderen Punkt  $p_j \in f$  durch einen endlichen Linienzug verbunden werden kann, so daß jeder Punkt des Linienzuges in  $f$  ist (Abb. 3.1).<sup>2</sup> Man beachte, daß die Definition der Fläche auch solche (gewissermaßen fiktiven) Flächen zuläßt, die nicht durch einen geschlossenen Kantenzug begrenzt sind. Diese Flächen bezeichnen wir als *ungebundene Flächen*. Wenn ein Knoten ein Endpunkt einer Kante ist, bezeichnen wir die Beziehung zwischen diesem Knoten und dieser Kante als *gebunden*. Ebenso sind sowohl eine Fläche und eine Kante auf der Grenze dieser Fläche als auch eine Fläche und ein Knoten auf der Grenze dieser Fläche gebunden. Der *Grad* eines Knotens  $v$  bezeichnet die Anzahl der an  $v$  gebundenen Kanten. Unter der *Komplexität* einer planaren Subdivision verstehen wir schließlich die Summe der Anzahl der Knoten, der Anzahl der Kanten und der Anzahl der Flächen.

Planare Subdivisionen sind typisch für kartographische Darstellungen und dienen seit jeher als geographische Beschreibungsmodelle räumlicher Verteilungen (in Form von Arealen, Regionen, Feldern usw.; vgl. Bartels 1968, 1970). Und mehr noch: Einer geographisch-physischen Welt, die aus sicht- und fühlbaren, körperlichen Dingen besteht, die sich auf der Erdoberfläche verteilen, sind planare Subdivisionen gewissermaßen inhärent: An dem Ort (sei er punkt-, linien- oder flächenhaft), an dem ein Ding existiert, kann zur selben Zeit kein anderes existieren, es sei denn, die beiden in Frage stehenden Dinge lassen sich thematisch voneinander unterscheiden: Wo eine Zuckerrübe wächst, kann keine zweite wachsen (höchstens direkt daneben), aber es kann dort sehr wohl auch Schwarzerde vorhanden sein (zumindest direkt unterhalb der Rübe; auf die Erdoberfläche projiziert fallen beide Orte zusammen). Die thematische Zuordnung von Mengen von Dingen in der Form, daß sich jede Menge einzeln in Form einer planaren Subdivision beschreiben bzw. kartographisch darstellen läßt, erlaubt so auf die einfachste Weise das Rasonieren über diese Dinge nach dem Koinzidenzprinzip (vgl. Kap. 1.2.1).

Es ist daher kein Wunder, daß planare Subdivisionen der Fläche auch für das methodische Instrumentarium Geographischer Informationssysteme von grundlegender Bedeutung sind. Eine der wichtigsten Aufgaben Geographischer Informationssysteme ist die Organisation von geometrischen Informationen (z.B. Flurstücken) und thematischen Informationen (z.B. Bodenart, Bodentyp, Bodennutzung etc.) in der Form, daß sowohl geometriebasierte Abfragen ('Was ist dort?') als auch themabasier-

<sup>2</sup>Einfacher ausgedrückt: Schneidet man aus einer Papiervorlage einer Flächenkarte eine Fläche aus, so mag der ausgeschnittene Schnipsel zwar Löcher aufweisen, er besteht aber immer aus genau einem Stück Papier und zerfällt niemals in mehrere Teile.



**Abbildung 3.2** Eine visuelle Verschneidung. Zwei durch mindestens zwei Punkte  $p_1$  und  $p_2$  in der Euklidischen Ebene referenzierte thematische Karten (links) können geometrisch korrekt übereinandergelegt und so ‘visuell verschneiden’ werden (rechts). Die Schnittflächen sind nun sichtbar, jedoch in ihren geometrischen Eigenschaften unbekannt.

te Abfragen (‘Wo ist *dies*?’) gleichermaßen beantwortet werden können. Arbeitet man mit modernen GIS-Implementationen, so läßt sich eine einfache geometriebasierte Abfrage lösen, indem man mit der Maus die entsprechende(n) Stelle(n) in der Bildschirmappe markiert und die thematischen Informationen in den daraufhin automatisch markierten Zeilen der zugeordneten thematischen Tabelle abliest (‘clickable map’). Umgekehrt läßt sich eine themabasierte Abfrage über die Selektion der abzufragenden Datenzeilen lösen, in Folge derer die zugeordneten geometrischen Strukturen in der Karte automatisch hervorgehoben werden. Neben solch einfachen Abfragen lassen sich auch komplexere geometriebasierte Probleme bearbeiten, die z.B. auf räumlichen Distanzen (‘Was ist *dort* in einem *Umkreis* mit bestimmtem Radius?’), Nachbarschaftsbeziehungen (‘Was ist in der *dortigen* Nachbarschaft [*i*-ten Grades?’) oder logischen Junktoren basieren (‘Was ist *hier* und/oder *dort*?’). Wie man leicht sieht, setzt die Lösung dieser und ähnlicher Probleme im Grunde genommen nur folgendes voraus:

1. einen Schlüssel (eine oder mehrere Zahlen bzw. Zeichenketten), über den die geometrischen und thematischen Informationen einander zugeordnet werden können,
2. eine (i.d.R. relationale) Datenbank, die die Daten vorhält und die Zuordnung widerspruchsfrei gewährleistet, sowie
3. diverse Datenstrukturen, mit Hilfe derer die geometrischen Informationen aufbereitet werden, um eine effektive (geometriebasierte) Suche zu gewährleisten.

Komplexe geometriebasierte Abfragen, mit denen aufgrund gegebener geometrischer Informationen Auskünfte über das Vorhandensein, die räumliche Verteilung, Verbreitung, Dichte usf. von thematischen Eigenschaften eingeholt werden, stellen neben den o.g. kaum weitere spezifische Anforderungen an ein Geographisches Informationssystem: Sind die Informationen einmal in geeigneter Form vorbearbeitet, lassen sich beliebige Abfragen effektiv und ohne weiteren Rechenaufwand durchführen. Die weitere Aufbereitung der über die geometrischen Eigenschaften abgefragten thematischen Informationen erfolgt dann nach arithmetischen Regeln und ist Aufgabe des Datenbanksystems bzw. anderer, nachgeschalteter Software.

Ganz anders verhält es sich dagegen mit komplexen themabasierten Abfragen (‘Wo ist *dies* und/oder *das*?’), die auf die räumliche Koinzidenz verschiedener Sachverhalte abzielen. Nehmen wir an, wir haben zwei digitale Karten: eine Karte der Flächen gleichen Bodentyps und eine der Flurstücke sowie zwei dazugehörige Tabellen mit den thematischen Attributen ‘Bodentyp’ und ‘Bodennutzung’. Neben

einfachen Abfragen ('Wo sind die Flächen mit Schwarzerde?' oder 'Wo werden Zuckerrüben angebaut?') sind häufig kombinierte Abfragen von Interesse ('Wo sind die Schwarzerdeflächen, auf denen Zuckerrüben angebaut werden?' bzw. 'Auf wieviel Prozent der Schwarzerdeflächen werden Zuckerrüben angebaut?'). Da die Flächen auf zwei digitalen Karten in der Regel nicht grenzscharf sind (d.h. die Flächengrenzen und -eckpunkte der einen Karte sind in der Regel nicht identisch mit denen der zweiten Karte – anderenfalls würden wir ja dieselbe Kartengrundlage benutzen), wird die Abarbeitung des logischen Junktors 'und' (Schwarzerde  $\wedge$  Zuckerrüben) nun zunächst zu einem geometrischen Problem: Gesucht sind die Schnittflächen aller Flächen der Bodentypenkarte mit dem Attribut 'Schwarzerde' und aller Flächen der Bodennutzungskarte mit dem Attribut 'Zuckerrüben'. Sind diese Schnittflächen einmal bekannt, ist deren Darstellung in Form einer neuen digitalen Karte (als Antwort auf die Frage nach dem Wo) sowie die Berechnung von deskriptiven Statistiken (als Antwort auf die Frage nach dem Flächenanteil) trivial (Abb. 3.2).

Im Unterschied zu geometriebasierten Abfragen erfordern komplexe themabasierte Abfragen also die Berechnung neuer geometrischer Strukturen. Diese sind aus den gegebenen geometrischen Informationen nicht unmittelbar ableitbar, da weder die Schnittpunkte der Flächengrenzen zweier übereinandergelegter Karten noch die Nachbarschaftsverhältnisse der resultierenden Schnittflächen von vorneherein bekannt sind. Das zugrundeliegende Problem wird unter Geographen als *Problem der Kartenverschneidung* (map overlay problem) angesprochen und ist in einer allgemeineren Notation als *Polygonschnitt Problem* (polygon intersection oder auch polygon clipping problem) bekannt (z.B. Aumann, Spitzmüller 1993, S. 155ff.; Ottmann, Widmayer 1993, S. 519ff.; de Berg et al. 1997, S. 19ff.; Mount 1997).

Hier sollen nun erdoberflächliche und abstrakte (Daten-)Räume gleichrangig betrachtet werden. Anders als auf der Erdoberfläche, wo keine zwei Objekte einer inhaltlichen Domäne an ein und demselben Ort vorkommen, können in abstrakten 'Datenräumen' die Orte von Objekten innerhalb ein und derselben thematischen Domäne zusammenfallen<sup>3</sup>. Mit anderen Worten: Räumliche Koinzidenz tritt dann nicht nur zwischen, sondern auch innerhalb von thematischen Domänen auf. Um diese Koinzidenzen geometrisch, d.h. – in Analogie zu den Zuckerrüben auf Schwarzerdeflächen – als Schnittflächen neu zu berechnen, fassen wir jedes geometrische Objekt als einzelne planare Subdivision der Ebene auf. In diesem Sinne geht es uns dann darum, einen Algorithmus zu entwerfen, mit dem endlich viele planare Subdivisionen simultan verschnitten werden können und wir als Ergebnis eine planare Subdivision erhalten, so daß für jede einzelne Teilfläche dieser Subdivision bekannt ist, aus welchen Originalflächen sie 'entstanden' sind (vgl. S. 76). Wir betrachten dabei ausschließlich sog. Vektormodelle, d.h. unsere Geometriedaten liegen in Form von Punkten, Linien und Flächen sowie symbolischen Informationen vor, die die topologischen Beziehungen der einzelnen geometrischen Objekte untereinander beschreiben.

In der geographischen Literatur werden algorithmische Details des Polygonschnitt- bzw. Kartenverschneidungsproblems nur am Rande erwähnt, und auch Geoinformatiker schrecken vor den methodischen Anforderungen einer Verschneidung von Vektormodellen zurück. So bemerkt Bartelme zu dem Problem, zwei kleinräumige Waldgebietskarten „mit vielen Verästelungen, Einbuchtungen und Waldlichtungen“ zu vereinigen: „Dieses Problem kann bei Vektormodellen beinahe unlösbar werden“, denn dieses und ähnliche Probleme erfordern „komplexe Schnitt- und Flächenberechnungsalgorithmen“ (1995, S. 51; vgl. Laurini, Thompson 1995, S. 280; Worboys 1995, S. 221). Die Geographie zeigt sich auch hier wieder als anwendungsorientierte Disziplin, die spezialisierte methodische Probleme lieber den Fachleuten überläßt.

Um einer Lösung näherzukommen, müssen wir uns der algorithmischen Geometrie (computational geometry) zuwenden. Die Verschneidung zweier Polygone ist vielfach untersucht und dokumentiert

<sup>3</sup>Vgl. z.B. Flächen gleicher Bodennutzung mit den Kreisflächen in Kap. 4.2

worden (zuletzt z.B. Vatti 1992). Vielfach werden dabei nur Spezialfälle betrachtet, z.B. ausschließlich konvexe (Sutherland, Hodgeman 1974; Foley, van Dam 1984; Foley et al. 1990; Laszlo 1995) oder hierarchisch verschachtelte Polygone (Didier, Grange 1995). Allgemeinere Ansätze berücksichtigen auch konkave Polygone und konkave Polygone mit Löchern (Weiler, Atherson 1974; Sechrest, Greenberg 1981; Liang, Barsky 1983; Montani, Re 1987; Malliot 1992; Vatti 1992; Glassner 1994a; de Berg et al. 1997; vgl. auch Franklin et al. 1994).

Die simultane Verschneidung von mehr als zwei beliebigen Polygonen in der Ebene (bzw. planaren Subdivisionen der Ebene) wird dagegen nur selten thematisiert. Hoffmann, Hopcroft, Karasick (1988) benennen immerhin einige fundamentale Probleme, die sich bei der simultanen Verschneidung von drei Polygonen ergeben (vgl. Kap. 3.5.2.2). Ottmann und Widmayer (1993) geben ein Verfahren an, mit Hilfe dessen für eine Menge von  $m$  ( $m > 2$ ) Polygonen alle Paare sich schneidender Polygone bestimmt werden können. Ein Verschneidealgorithmus findet sich jedoch nicht. Erschwerend kommt hinzu, daß bei einer simultanen Verschneidung von beliebig vielen, i.d.R. eng benachbarten Polygonen die 'klassischen' Verschneideoperationen aufgrund ungenauer Arithmetik versagen (vgl. Kap. 3.5). Neben dem Problem, einen gegebenen Verschneidealgorithmus auf die Verschneidung von  $m$  Polygonen zu verallgemeinern, besteht die Aufgabe daher insbesondere auch darin, ein robustes, d.h. ein unabhängig vom numerischen Input stabiles und korrektes Verfahren anzugeben. Auch für dieses zweite Problemfeld existieren in erster Linie Untersuchungen von Teil- bzw. Spezialproblemen, z.B. über einfache Linienschnittprobleme (Walton, Xu 1991; Manocha, Demmel 1994) oder die dreifache Überschneidung von Liniensegmenten (Barker 1995). Ein Algorithmus, mit dem beliebig viele Polygone (konvexe und konkave, mit und ohne Löcher oder Inseln) bzw. planare Subdivisionen verschnitten werden können und der für jeden noch so kleinräumigen Input ein valides Ergebnis erzeugt, steht jedoch noch aus.

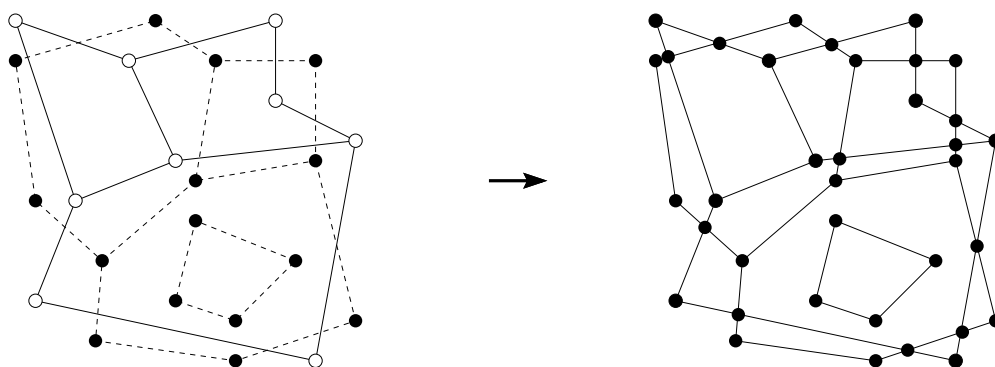
Im folgenden wird zunächst das Verfahren der Verschneidung zweier Flächenkarten beschrieben, wie es häufig mit GIS implementiert ist. Die nachstehenden Ausführungen in Kap. 3.2 halten sich eng an die Darstellung von de Berg et al. (1997, 19ff.), und auch die formalen und graphischen Notationen dieses sehr empfehlenswerten Werks werden übernommen. Alle abweichenden Schritte werden an entsprechender Stelle vermerkt.

## 3.2 Die Verschneidung zweier Flächenkarten

Wir betrachten im folgenden ausschließlich zweidimensionale Flächenkarten (d.h. Teilmengen der Euklidischen Ebene) und gehen ohne Einschränkung davon aus, daß die beiden zu verschneidenden Flächenkarten jeweils als planare Subdivision vorliegen (Abb. 3.1). Das zu lösende Problem formulieren wir zunächst als

**Problem der Kartenverschneidung:** Die Verschneidung zweier planarer Subdivisionen  $\mathcal{S}_1$  und  $\mathcal{S}_2$  ist die planare Subdivision  $Sub(\mathcal{S}_1, \mathcal{S}_2)$ , für die eine Fläche  $f$  genau dann existiert, wenn es eine Fläche  $f_1$  in  $\mathcal{S}_1$  und eine Fläche  $f_2$  in  $\mathcal{S}_2$  gibt, so daß  $f$  eine maximal verbundene Teilmenge aus  $f_1 \cap f_2$  ist. Bestimme  $Sub(\mathcal{S}_1, \mathcal{S}_2)$ .

Mit anderen Worten: Das allgemeine Kartenverschneidungsproblem zu lösen bedeutet im wesentlichen, diejenige planare Subdivision der Ebene zu finden, die durch die Kanten aus  $\mathcal{S}_1$  und  $\mathcal{S}_2$  induziert wird (Abb. 3.3). Das Gesamtproblem läßt sich nun in unterschiedliche Teilprobleme gliedern, die im folgenden getrennt voneinander behandelt werden:



**Abbildung 3.3** Eine geometrische Situation vor und nach der Verschneidung. Eine Verschneidung zu berechnen bedeutet im wesentlichen, den planaren Graphen zu bestimmen, der durch die Knoten und Kanten einer gegebenen geometrischen Situation induziert wird. Dazu müssen sämtliche Schnittpunkte der Kanten bestimmt werden.

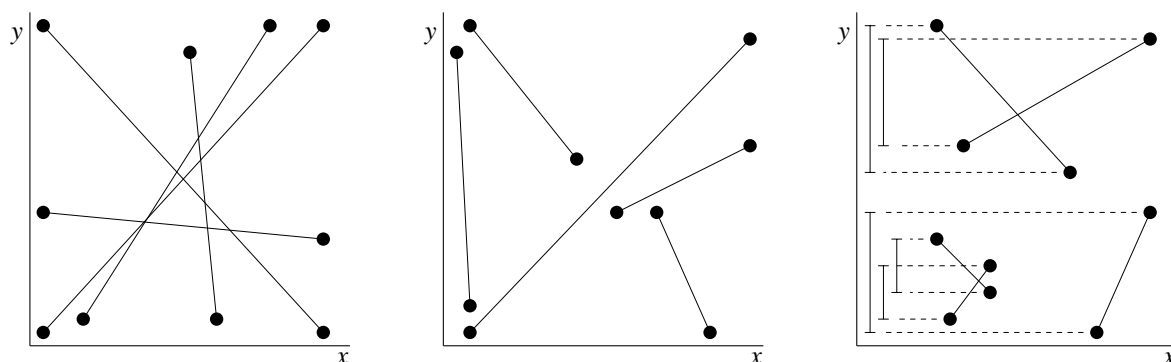
- Das *geometrische* Teilproblem beinhaltet die Berechnung der Schnittpunkte aller Kanten aus  $\mathcal{S}_1$ ,  $\mathcal{S}_2$ .
- Das *topologische* Teilproblem beinhaltet die Konstruktion des der Subdivision  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  zugrundeliegenden planaren Graphen, die Konstruktion der neuen (Teil-)Flächen  $f \in Sub(\mathcal{S}_1, \mathcal{S}_2)$  aus den sie umschließenden Flächengrenzen (den Zyklen dieses Graphen) sowie die Zuordnung jeder Fläche  $f$  zu den beiden Flächen  $f_1 \in \mathcal{S}_1$  und  $f_2 \in \mathcal{S}_2$ , deren Schnittfläche  $f$  darstellt.
- Das *thematische* Teilproblem beinhaltet schließlich die Neuberechnung der thematischen Informationen für jede Fläche  $f = f_1 \cap f_2 \in Sub(\mathcal{S}_1, \mathcal{S}_2)$  anhand der thematischen Informationen für  $f_1 \in \mathcal{S}_1$  und  $f_2 \in \mathcal{S}_2$  und gegebener Boolescher und/oder arithmetischer Regeln.

### 3.2.1 Der Scan-Line-Algorithmus

Wie man sieht, ist die einzige im engen Sinne geometrische Operation die Berechnung von Schnittpunkten der Kanten aus  $\mathcal{S}_1$  und  $\mathcal{S}_2$ , alle (!) anderen Operationen sind dagegen symbolischer Natur. Wie aber finden wir nun alle Schnittpunkte? In einem ersten Lösungsansatz könnte man jede Kante aus  $\mathcal{S}_1$  mit jeder Kante aus  $\mathcal{S}_2$  auf einen Schnittpunkt hin untersuchen und diesen berichten, wenn er existiert. Ein solcher, naiver Algorithmus erfordert für  $m$  Kanten in  $\mathcal{S}_1$  und  $n$  Kanten in  $\mathcal{S}_2$  genau  $m \cdot n$  Schnitpunkttests, und zwar unabhängig davon, ob sich keine zwei Kanten oder alle Kanten paarweise überschneiden. Der Einfachheit halber vernachlässigen wir für einen Moment die Existenz zweier planarer Subdivisionen und betrachten zunächst ein

**Allgemeines Segment-Schnitt-Problem:** Sei  $S := \{s_1, s_2, \dots, s_n\}$  eine Menge von Liniensegmenten, jedes Liniensegment  $s_i$  sei durch je zwei Punkte  $p_l$  und  $p_u$  der Euklidischen Ebene gegeben,  $s_i = \{p_u, p_l\}$ ,  $1 \leq i \leq n$ , und zwei beliebige Liniensegmente schneiden sich in höchstens einem Punkt, d.h. sie überlappen sich nicht. Bestimme alle Schnittpunkte  $p_{ij} = s_i \cap s_j$ ,  $1 \leq i < j \leq n$ .

Man beachte, daß die Existenz eines sog. *degenerierten* Falls bei den folgenden Erörterungen zunächst ausgeschlossen wird, und zwar die Überlappung zweier oder mehrerer Kanten (vgl. dazu Kap. 3.4). Der oben skizzierte naive Algorithmus erfordert  $\frac{n^2-n}{2}$  Schnitpunkttests, die Anzahl der durchgeführten Schnitpunkttests steigt also quadratisch mit der Anzahl der Liniensegmente. Wir schreiben daher: Der Algorithmus benötigt  $O(n^2)$  Zeit, bzw. ist in der Ordnung  $O(n^2)$  (zur Analyse der Effizienz von



**Abbildung 3.4** Verschiedene Anordnungen von Liniensegmenten. Mehrere Liniensegmente können sich alle paarweise (links) oder auch gar nicht überschneiden (Mitte). Wenn sich zwei Linien schneiden, überlappen sich deren  $y$ -Intervalle (rechts).

Algorithmen vgl. ausführlich Brassard, Bratley 1993, S. 22ff.). Unter gewissen Umständen ist diese Zeit optimal, dann nämlich, wenn sich jedes Segment mit jedem überschneidet (Abb. 3.4 links). In der Praxis stellt diese Situation jedoch die seltene Ausnahme dar, denn in der Regel überschneiden sich die meisten Segmente mit keinem oder nur wenigen anderen Segmenten, so daß die Anzahl der existierenden Schnittpunkte wesentlich kleiner als quadratisch ist. In diesen Fällen sollten wir von unserem Algorithmus verlangen, daß er auch weniger als quadratische Zeit benötigt. Die Laufzeit des gesuchten Verfahrens soll also nicht nur von der Größe des Inputs (hier:  $n$  Segmente), sondern auch von der Größe des Outputs (hier: Anzahl der Schnittpunkte) abhängig sein. Mit anderen Worten: Das gesuchte Verfahren ist Output-sensitiv.

Um zu vermeiden, jedes Segmentpaar auf einen Schnittpunkt zu testen, müssen wir uns die vorliegende geometrische Situation zunutze machen; Segmente, die nah beieinander liegen, sind offensichtlich eher Kandidaten mit einem gemeinsamen Schnittpunkt als solche, die weit voneinander entfernt sind. Eine einfache Möglichkeit, die Menge der zu testenden Segmentpaare einzuschränken, besteht nun z.B. darin, die vertikale Lage der Segmente zu berücksichtigen: Sei das  $y$ -Intervall eines Segmentes die orthogonale Projektion des Segmentes auf die  $y$ -Achse. Wenn sich die  $y$ -Intervalle zweier Segmente nicht überlappen (wir könnten in diesem Fall sagen, daß die beiden Segmente auf der Vertikalen weit voneinander ‘entfernt’ sind), dann haben die beiden Segmente mit Sicherheit keinen Punkt gemeinsam. Wir müssen daher nur jene Segmentpaare auf einen Schnittpunkt hin untersuchen, deren  $y$ -Intervalle sich überschneiden, für die also eine horizontale Linie existiert, die beide Segmente schneidet (Abb. 3.4 rechts). Um diese Paare zu finden, stellen wir uns eine horizontale Linie  $\ell$  vor, die, von einer Position oberhalb der Segmente ausgehend, von oben nach unten über die Ebene geschwenkt wird. Während des Schwenks dieser imaginären Horizontalen achten wir peinlich genau auf alle Segmente, die gerade von ihr geschnitten werden, und finden so die gesuchten Paare.

Ein solcher Typ von Algorithmus ist unter den Bezeichnungen *plane-sweep-algorithm* (de Berg et al. 1997; Mount 1997) oder *Scanline-Prinzip* (Aumann, Spitzmüller 1993, S. 162ff.; Ottmann, Widmayer 1993, S. 464ff.; Antonio 1994; Prasad 1994; de Berg et al. 1997, S. 22ff.) bekannt, die gedachte Horizontale  $\ell$  wird im deutschen Sprachgebrauch als Scanlinie bezeichnet. Unter dem Status der Scanlinie  $\ell$  verstehen wir die Menge der Segmente, die von  $\ell$  geschnitten werden. Der Status verändert sich während des Schwenks, allerdings nicht kontinuierlich. Eine Aktualisierung des Status ist nämlich nur an einzelnen Punkten erforderlich, den sog. Eventpunkten. Diese Eventpunkte sind zunächst die Endpunkte der Segmente, während des Scanlineschwenks werden die gefundenen Schnittpunkte zur Menge der Eventpunkte hinzugefügt. Die Scanlinie wird, beginnend beim obersten Eventpunkt,



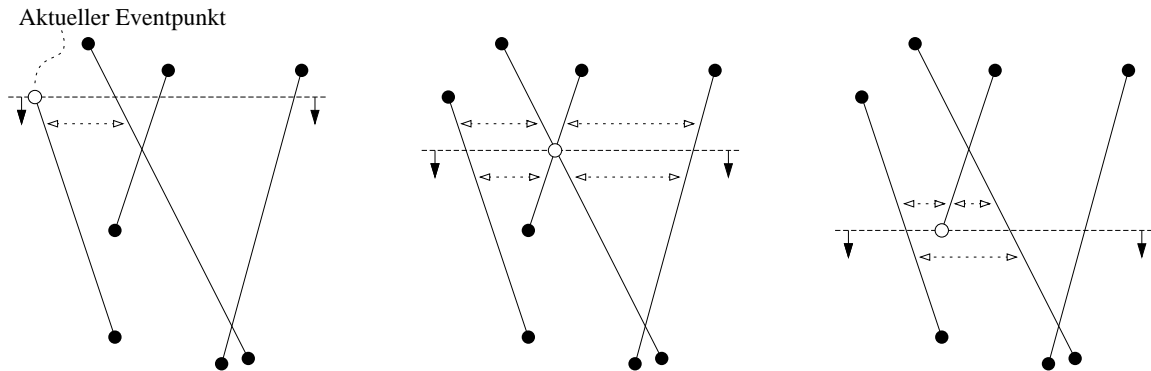
von Eventpunkt zu Eventpunkt nach unten versetzt, und das Verfahren terminiert, wenn der unterste Eventpunkt erreicht ist. An jedem neuen Eventpunkt, den die Scanlinie erreicht, müssen der Status aktualisiert und – dem Status entsprechend – Schnittpunkttests durchgeführt werden: Jene Segmente, deren obere Endpunkte jetzt auf der Scanlinie liegen, müssen zum Status hinzugefügt werden, und jene Segmente, deren untere Endpunkte jetzt auf der Scanlinie liegen, müssen vom Status entfernt werden. Nur die Segmente, die dann zum Status gehören (die also von der momentanen Scanlinie geschnitten werden und deren unterer Endpunkt unterhalb der Scanlinie liegt), müssen anschließend auf Intersektionen geprüft werden.<sup>4</sup>

Um die Zahl der Schnittpunkttests weiter zu reduzieren, wird der Status von links nach rechts sortiert. Es genügt dann, jedes neu in den Status eingefügte Segment mit seinem direkten linken und seinem direkten rechten Nachbarn auf Intersektion zu prüfen. Später, wenn die Scanlinie weiter nach unten geschwenkt ist, mag es für dieses Segment andere Nachbarn geben, mit denen ein Schnittpunkttest durchgeführt wird. Diese Strategie muß natürlich im Status entsprechend reflektiert werden: Unter dem Status verstehen wir im folgenden die in der Horizontalen von links nach rechts geordnete Sequenz aller Segmente, die von der Scanlinie geschnitten werden. Für den Fall, daß sich zwei Segmente genau auf der Scanlinie schneiden, ist die Reihenfolge direkt unterhalb der Scanlinie ausschlaggebend. Der Status verändert sich jetzt nicht nur an den Endpunkten der Segmente, sondern auch an den im laufenden Verfahren gefundenen Schnittpunkten. Die Schnittpunkte stellen damit einen neuen Typ von Eventpunkten dar, an denen sich die Ordnung der sich schneidenden Segmente ändert, so daß diese auf Intersektion mit ihren neuen linken und rechten Nachbarn geprüft werden müssen (Abb. 3.5). Im einzelnen unterscheiden wir also drei Fälle:

- Wenn der aktuelle Eventpunkt der obere Endpunkt eines Segmentes  $s$  ist, ist  $s$  ein neues Segment, das die Scanlinie schneidet. Daher wird  $s$  zum Status hinzugefügt und auf Intersektion mit seinem linken und seinem rechten Nachbarn geprüft. Wird ein Schnittpunkt unterhalb der Scanlinie entdeckt, wird dieser zu einem neuen Eventpunkt. Nachdem der aktuelle Eventpunkt bearbeitet worden ist (vgl. Kap. 3.2.3), fahren wir am nächsten Eventpunkt fort (Abb. 3.5 links).
- Wenn der aktuelle Eventpunkt ein Schnittpunkt ist, müssen die beiden sich schneidenden Segmente ihre Reihenfolge im Status tauschen. Der Grund dafür ist leicht zu sehen: Wenn zwei Segmente  $s_i$  und  $s_j$  sich in einem gemeinsamen Punkt  $p$  schneiden, so daß  $p$  nicht ein Endpunkt von  $s_i$  oder  $s_j$  ist, und oberhalb von  $p$  die Relation ‘ $s_i$  ist links von  $s_j$ ’ gilt, dann gilt unterhalb von  $p$  immer die Relation ‘ $s_i$  ist rechts von  $s_j$ ’. Für jedes der beiden Segmente existiert dann (höchstens) ein neuer Nachbar, der auf Intersektion geprüft werden muß. Wiederum wird jeder neu entdeckte Schnittpunkt unterhalb der Scanlinie zu einem neuen Eventpunkt (Abb. 3.5 Mitte).
- Wenn der aktuelle Eventpunkt der untere Endpunkt eines Segmentes  $s$  ist, werden der linke Nachbar  $s_l$  und der rechte Nachbar  $s_r$  von  $s$  unterhalb der Scanlinie zu direkten Nachbarn. Das Segment  $s$  wird vom Status entfernt und der Schnittpunkt von  $s_l$  und  $s_r$  – falls er existiert – zu einem neuen Eventpunkt (Abb. 3.5 rechts).

Nachdem das Prinzip des Verfahrens weitestgehend geklärt ist, wenden wir uns nun den verfahrenstechnischen Details zu, soweit dies für die Formulierung eines allgemeinen Algorithmus erforderlich ist. Zunächst benötigen wir eine Datenstruktur für die Speicherung der einzelnen Eventpunkte. Wir nennen diese Datenstruktur *Eventfolge* und bezeichnen sie mit  $Q$ . Von der Eventfolge erwarten wir,

<sup>4</sup>Es sind Situationen vorstellbar, in denen kein einziger Schnittpunkt zwischen den Segmenten existiert, und trotzdem eine quadratische Anzahl an Schnittpunkttests durchgeführt werden muß, in dem Fall nämlich, wenn sich die  $y$ -Intervalle der Segmente alle paarweise überlappen, die Segmente in der Horizontalen jedoch genügend weit voneinander entfernt sind.



**Abbildung 3.5** Drei aufeinanderfolgende Eventpunkte eines 'Scanline-Schwenks'. Die Nachbarschaftsbeziehungen, die sich am aktuellen Eventpunkt zwischen den Segmenten auf der Scanlinie ändern, sind jeweils vor Bearbeitung (oberhalb der Scanlinie) und nach Bearbeitung (unterhalb der Scanlinie) eines Eventpunktes dargestellt.

daß sie die effektive Suche sowie das effektive Einfügen und Löschen von einzelnen Eventpunkten erlaubt. Zusammen mit jedem Eventpunkt  $p$  in  $\mathcal{Q}$  werden all jene Segmente gespeichert, deren oberer Endpunkt  $p$  ist.<sup>5</sup> Da die Endpunkte verschiedener Segmente aufeinander fallen können (im Falle planarer Subdivisionen ist dies die Regel!), muß bei einer Einfügeoperation überprüft werden können, ob ein Eventpunkt bereits in  $\mathcal{Q}$  existiert. Insbesondere muß für einen aktuellen Eventpunkt der nächst folgende Eventpunkt festgestellt werden können (das ist der oberste Eventpunkt unterhalb der Scanlinie). Die Reihenfolge soll im Sinne unserer Scanlinie von oben nach unten verlaufen, und für den Fall, daß zwei Eventpunkte dieselbe  $y$ -Koordinate haben, legen wir fest, daß der Eventpunkt mit der kleineren  $x$ -Koordinate zuerst bearbeitet wird. Die Abarbeitung der Eventpunkte soll somit von oben nach unten und – bei gleicher Höhe – von links nach rechts verlaufen. Etwas formaler ausgedrückt definieren wir eine Ordnung  $\prec_{\mathcal{Q}}$  auf den Eventpunkten, die die Reihenfolge repräsentiert, in der die Eventpunkte abgearbeitet werden sollen. Für zwei Eventpunkte  $p$  und  $q$  gilt also  $p \prec_{\mathcal{Q}} q$  (in Worten:  $p$  ist in  $\mathcal{Q}$  Vorgänger von  $q$ ) dann und nur dann, wenn entweder  $p.y > q.y$  oder wenn  $p.y = q.y$  und  $p.x < q.x$ . Naheliegenderweise wird die Eventfolge als ein balancierter Binärbaum implementiert, dessen Sortierkriterium der Ordnung  $\prec_{\mathcal{Q}}$  entspricht. Sämtliche Such-, Einfüge- und Löschooperationen sind so in  $O(\log n)$  möglich (vgl. Aumann, Spitzmüller 1993, S. 155ff.; Ottmann, Widmayer 1993, S. 297ff.).

Zusätzlich zur Eventfolge benötigen wir eine Datenstruktur, die es uns ermöglicht, den Status des Verfahrens (die in der Horizontalen von links nach rechts geordnete Sequenz aller Segmente, die von der Scanlinie geschnitten werden; vgl. oben) zu überwachen. Wir nennen diese Datenstruktur *Statusfolge* und bezeichnen sie mit  $\mathcal{T}$ . Wir definieren eine Ordnung  $\prec_{\mathcal{T}}$  auf den Segmenten des Status, die die Reihenfolge repräsentiert, in der die Segmente von der Scanlinie geschnitten werden. Für zwei Segmente  $s_i$  und  $s_j$  gilt  $s_i \prec_{\mathcal{T}} s_j$  dann und nur dann, wenn entweder zwei Punkte  $p_i = s_i \cap \ell$  und  $p_j = s_j \cap \ell$  existieren, so daß  $p_i.x < p_j.x$  ist oder wenn  $p_i.x = p_j.x$  ist und zwei Punkte  $\hat{p}_i = s_i \cap \hat{\ell}$  und  $\hat{p}_j = s_j \cap \hat{\ell}$  existieren, so daß  $\hat{p}_i.x < \hat{p}_j.x$  ist, wobei  $\hat{\ell}$  eine horizontale Linie kurz unterhalb von  $\ell$  ist, so daß weder auf  $\hat{\ell}$  noch zwischen  $\hat{\ell}$  und  $\ell$  ein Eventpunkt liegt. Mit anderen

<sup>5</sup>Konkrete Implementierungen, die eine solche 'Bündelung' von Daten erlauben, sind z.B. Strukturen (wie in ANSI C) oder Objekte (wie in objektorientierten Programmiersprachen, etwa C++ oder Objective C). Sie spielen für das Verständnis des Verfahrens jedoch keine Rolle. Es genügt die Vorstellung, daß 'gebündelte' Daten direkt und ohne Suche verfügbar sind. Hier bedeutet es, daß bei Kenntnis des Eventpunktes  $p$  alle Segmente, deren oberer Endpunkt  $p$  ist, ebenfalls bekannt sind. Auf die gleiche Weise werden im übrigen auch zwei Koordinaten zu einem Punkt und zwei Punkte zu einer Linie 'gebündelt'.

Worten: Entscheidend für die Relation  $s_i$  ‘ist links (rechts) von’  $s_j$  ist entweder die Lage der Punkte, in denen die beiden Segmente die Scanlinie schneiden, oder – für den Fall, daß der Schnittpunkt von  $s_i$  und  $s_j$  genau auf der Scanlinie liegt – die Lage der Punkte, in denen die beiden Segmente eine Linie kurz unterhalb der Scanlinie schneiden. Der Fall, daß ein oder mehrere Segmente mit ihren unteren Endpunkten auf der Scanlinie liegen, ist ausgeschlossen, denn sie würden, bevor die Scanlinie diese Position erreicht, von der Statusfolge entfernt. Horizontale Segmente hingegen können auftreten (sie liegen dann vollständig auf der Scanlinie), für die wir – analog zur Ordnung der Eventfolge – den linken Endpunkt als oberen Endpunkt festlegen. In der von links nach rechts sortierten Folge von Segmenten mit einem gemeinsamen Punkt  $p \in \ell$  steht ein horizontales Segment immer am Ende (mehrere horizontale Segmente mit einem gemeinsamen Punkt  $p \in \ell$  treten wiederum nicht auf, da wir die Überlappung von Segmenten für den Moment ausgeschlossen haben; auch die Berührung zweier horizontaler Segmente in  $p \in \ell$  ist ausgeschlossen, da in diesem Fall  $p$  für eines der Segmente der untere Endpunkt wäre).

Auch die Statusfolge muß effektive Such-, Einfüge- und Löschoptionen erlauben, und da die Reihenfolge der Segmente des Status eindeutig ist, wird die Statusfolge ebenfalls als balancierter Binärbaum implementiert, dessen Sortierkriterium  $\prec_{\mathcal{T}}$  entspricht. Man beachte, daß die Ordnung  $\prec_{\mathcal{T}}$  der Statusfolge (im Gegensatz zur Ordnung der Eventfolge!) eine lokale Ordnung ist, d.h., sie ändert sich im Laufe des Verfahrens, und zwar in Abhängigkeit vom aktuellen Eventpunkt: Für zwei sich schneidende Segmente  $s_i$  und  $s_j$ , für die oberhalb ihres gemeinsamen Schnittpunkts  $p$  (der nicht ein Endpunkt von  $s_i$  oder  $s_j$  ist)  $s_i \prec_{\mathcal{T}} s_j$  gilt, gilt immer auch  $s_j \prec_{\mathcal{T}} s_i$  unterhalb von  $p$  (vgl. oben). Ein solcher Wechsel der Reihenfolge muß natürlich in der Statusfolge reflektiert werden. Dies erreicht man am einfachsten, indem man die beiden Segmente zunächst von der Statusfolge entfernt, der Statusfolge die neue lokale Ordnung (als neues Sortierkriterium) zuweist und anschließend die Segmente wieder einfügt.

Die Eventfolge  $\mathcal{Q}$  und die Statusfolge  $\mathcal{T}$  sind die beiden einzigen Datenstrukturen, die im Moment benötigt werden. Das ‘Gerüst’ des gesuchten Algorithmus ist nun wie folgt:

**Algorithmus SCHNITTPUNKTSUCHE( $S$ )**

*Eingabe:* Eine Menge  $S$  von Liniensegmenten in der Euklidischen Ebene.

*Ausgabe:* Die Menge der Schnittpunkte zwischen allen Segmenten in  $S$  und für jeden Schnittpunkt die Menge der Segmente, die ihn enthalten.

1. Initialisiere eine leere Eventfolge  $\mathcal{Q}$  und füge anschließend die Endpunkte aller Segmente aus  $S$  in  $\mathcal{Q}$  ein. Wird ein oberer Segmentpunkt eingefügt, muß mit dem Eventpunkt auch das entsprechende Segment gespeichert werden.
  2. Initialisiere eine leere Statusfolge  $\mathcal{T}$ .
  3. Bestimme den aktuellen Eventpunkt:  $p :=$  der erste Eventpunkt in  $\mathcal{Q}$ .
  4. **while** ( $p$  existiert)
  5.     BEARBEITEEVENTPUNKT( $p$ )
  6.      $p :=$  der Nachfolger von  $p$  in  $\mathcal{Q}$ .
- end while**

Dieser Algorithmus gibt zunächst nur die Abarbeitung der einzelnen Eventpunkte in der mit  $\mathcal{Q}$  gegebenen Reihenfolge vor.<sup>6</sup> Wir haben bereits gesehen, was an jedem einzelnen Eventpunkt  $p$  geschehen muß: An Endpunkten von Segmenten müssen Segmente in die Statusfolge  $\mathcal{T}$  eingefügt bzw. daraus gelöscht werden, an Schnittpunkten muß die Reihenfolge von (mindestens) zwei Segmenten vertauscht

<sup>6</sup>de Berg et al. (1997, S. 25) schlagen vor, die Eventfolge  $\mathcal{Q}$  bereits im laufenden Verfahren zu leeren, indem der aktuelle Eventpunkt nach seiner Bearbeitung jeweils aus  $\mathcal{Q}$  gelöscht wird. Der nächst folgende aktuelle Eventpunkt ist dann immer das erste Element in  $\mathcal{Q}$ , und das Verfahren terminiert, wenn  $\mathcal{Q}$  leer ist. Die Ausführungen in Kap. 3.5 werden zeigen, daß wir die Eventfolge  $\mathcal{Q}$  unter bestimmten Voraussetzungen mehrfach durchlaufen müssen, weshalb sie an dieser Stelle auch nicht geleert werden darf. Auf die Laufzeit des Verfahrens wirkt sich die Änderung nicht aus.

werden. In beiden Fällen müssen für die Segmente, die am Eventpunkt  $p$  in  $\mathcal{T}$  direkt benachbart sind, Schnittpunkttests durchgeführt und gefundene Schnittpunkte in die Eventfolge  $\mathcal{Q}$  eingefügt werden. Die nächste Prozedur beschreibt dieses Vorgehen genau:

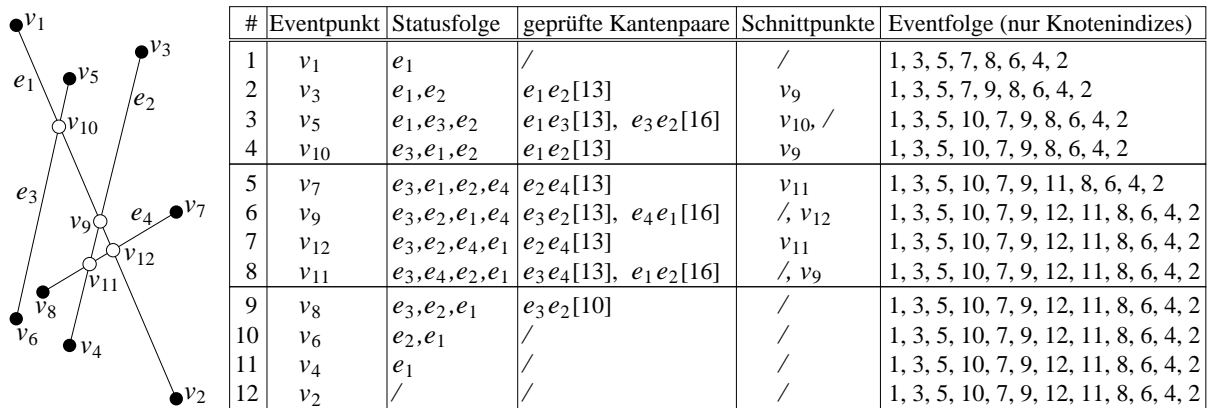
**Algorithmus** BEARBEITEEVENTPUNKT( $p$ )

*Eingabe:* Ein Eventpunkt  $p$ .

*Ausgabe:* Der Eventpunkt  $p$ , wenn  $p$  ein Schnittpunkt ist.

1. Sei  $U(p)$  die Menge der Segmente, deren oberer Endpunkt  $p$  ist. (Diese Segmente sind mit dem Eventpunkt  $p$  gespeichert. Der obere Endpunkt eines horizontalen Segmentes ist per definitionem der linke Endpunkt.)
2. Suche in  $\mathcal{T}$  nach der Menge  $S(p)$  aller Segmente, die  $p$  enthalten (sie sind in  $\mathcal{T}$  benachbart). Sei  $L(p) \subset S(p)$  die Menge der Segmente, deren unterer Endpunkt  $p$  ist, und  $C(p) \subset S(p)$  die Menge der Segmente, die  $p$  enthalten und von denen  $p$  nicht ein Endpunkt ist.
3. **if**  $(U(p) \cup C(p) \cup L(p)$  enthält mehr als ein Segment)  
**then**
4.     Berichte  $p$  als Schnittpunkt, zusammen mit  $U(p) \cup C(p) \cup L(p)$ .  
**end if**
5. Lösche die Segmente  $C(p) \cup L(p)$  aus  $\mathcal{T}$
6. Weise  $\mathcal{T}$  die neue lokale Ordnung für  $p$  zu
7. Füge die Segmente  $U(p) \cup C(p)$  in  $\mathcal{T}$  ein. (Löschen und Wiedereinfügen der Segmente aus  $C(p)$  vertauscht ihre Reihenfolge; gibt es ein horizontales Segment in  $U(p) \cup C(p)$ , so steht es von diesen an letzter Stelle. Die Scanlinie verläuft nun durch  $p$ , und die Ordnung aller Segmente in  $\mathcal{T}$  entspricht der Reihenfolge, in der sie von der Scanlinie geschnitten werden.)
8. **if**  $(U(p) \cup C(p) = \emptyset)$   
**then**
9.     Seien  $s_l$  und  $s_r$  der linke und rechte Nachbar von  $p$  in  $\mathcal{T}$ .
10.     FINDESCHNITTPUNKTEVENT( $s_l, s_r, p$ )  
**else**
11.     Sei  $s_{ll}$  das am weitesten links stehende Segment aus  $U(p) \cup C(p)$  in  $\mathcal{T}$ .
12.     Sei  $s_l$  der linke Nachbar von  $s_{ll}$  in  $\mathcal{T}$ .
13.     FINDESCHNITTPUNKTEVENT( $s_l, s_{ll}, p$ )
14.     Sei  $s_{rr}$  das am weitesten rechts stehende Segment aus  $U(p) \cup C(p)$  in  $\mathcal{T}$ .
15.     Sei  $s_r$  der rechte Nachbar von  $s_{rr}$  in  $\mathcal{T}$ .
16.     FINDESCHNITTPUNKTEVENT( $s_r, s_{rr}, p$ )  
**end if**

Diese Schnittpunktssuche selbst ist einfach. Zwei Segmente werden auf Intersektion getestet, und wenn ein Schnittpunkt existiert, wird dieser in die Eventfolge  $\mathcal{Q}$  eingefügt (es sei denn, er ist bereits in  $\mathcal{Q}$ ). Wir müssen allerdings darauf achten, daß wir nur solche Schnittpunkte berücksichtigen, die nicht bereits an früherer Stelle als aktueller Eventpunkt bearbeitet worden sind. Sind keine horizontalen Segmente vorhanden, dann liegt jeder gefundene Schnittpunkt in jedem Fall unterhalb der Scanlinie (d.h. auch unterhalb des aktuellen Eventpunktes) und ist also ein neuer, zuvor unbekannter Eventpunkt. Gibt es dagegen horizontale Segmente, müssen wir uns an unsere Konvention halten, daß Eventpunkte mit derselben  $y$ -Koordinate von links nach rechts bearbeitet werden. Daher sind wir sehr wohl auch an solchen Schnittpunkten interessiert, die wir auf der Scanlinie rechts vom aktuellen Eventpunkt finden. Mit anderen Worten: Wir betrachten nur solche Schnittpunkte als neue Eventpunkte, die noch nicht in  $\mathcal{Q}$  sind und, der Ordnung  $\prec_{\mathcal{Q}}$  der Eventfolge entsprechend, unterhalb oder auf gleicher Höhe und rechts vom aktuellen Eventpunkt liegen. Das Verfahren zur Bestimmung eines Schnittpunktes und der Aktualisierung der Eventfolge ist daher wie folgt:



**Abbildung 3.6** Ein Segmentschnittproblem und seine Lösung mit Hilfe des Scanline-Verfahrens. Die Angaben beziehen sich jeweils auf den Zustand *nach* der Bearbeitung des aktuellen Eventpunktes. Weiß markierte Knoten sind Schnittpunkte, die im Laufe des Verfahrens gefunden werden. Die Zusätze in eckigen Klammern bezeichnen die Zeile in BEARBEITEEVENTPUNKT, nach der ein Kantenpaar auf Intersektion geprüft wird.

**Algorithmus** FINDESCHNITTPUNKTEVENT( $s_l, s_r, p$ )

Eingabe: Zwei Segmente  $s_l$  und  $s_r$  sowie der aktuelle Eventpunkt  $p$ .

Ausgabe: Keine.

1. **if** ( $s_l$  und  $s_r$  existieren und  $p_i = s_l \cap s_r \neq \emptyset$  und  $p \prec_Q p_i$  und  $p_i \notin Q$ )  
**then**
2.     Füge  $p_i$  als neuen Eventpunkt in  $Q$  ein.
- end if**

Damit ist das Verfahren zur Lösung des Allgemeinen Segment-Schnitt-Problems vollständig beschrieben. Wir sehen nun auch leicht eine *Invariante des Scan-Line-Prinzips*, nach der ausnahmslos alle Schnittpunkte oberhalb der Scanlinie  $\ell$  korrekt berechnet worden sind. Diese Invariante gilt an jeder beliebigen Position der Scanlinie  $\ell$  und ist von zentraler Bedeutung für das Verständnis des allgemeinen Verfahrens und insbesondere der Behandlung von Rundungsfehlern, so wie sie im folgenden Kapitel besprochen wird. Ein einfaches Beispiel zeigt Abb. 3.6.

de Berg et al. (1997, S. 23 u. S. 27-29) zeigen, daß mit dem Verfahren ausnahmslos alle Schnittpunkte gefunden werden und daß der Algorithmus Output-sensitiv ist:

**Satz 3.1.** Sei  $S$  eine Menge von  $n$  Liniensegmenten in der Euklidischen Ebene. Das Problem, sämtliche Schnittpunkte zwischen den Segmenten in  $S$  zu berechnen und für jeden Schnittpunkt die Segmente, die ihn enthalten, zu berichten, kann in  $O(n \log n + I \log n)$  Zeit gelöst werden, wobei  $I$  die Anzahl der Schnittpunkte ist.

**3.2.2 Die doppelt-verkettete Kantenliste**

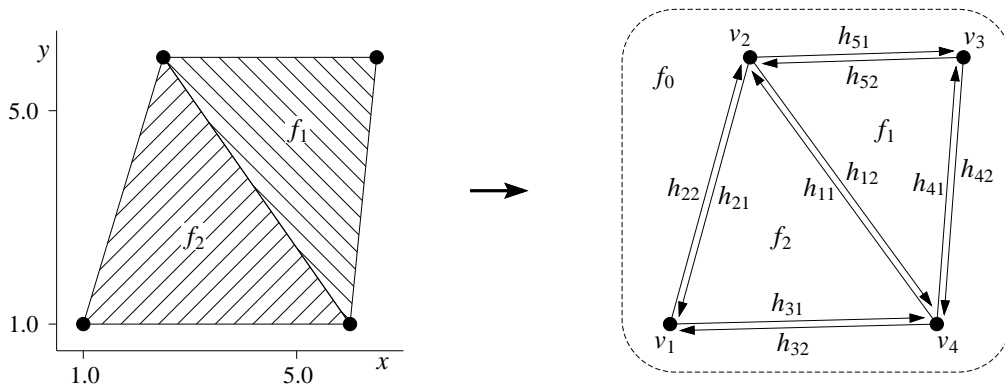
Kommen wir nun auf unser eigentliches Problem zurück, wie das Overlay zweier planarer Subdivisionen  $S_1$  und  $S_2$  als planare Subdivision  $Sub(S_1, S_2)$  bestimmt wird. Ein wichtiges (das geometrische; vgl. S. 55) Teilproblem ist bereits gelöst, denn zur Bestimmung aller Schnittpunkte der Flächengrenzen aus  $S_1 \cup S_2$  können wir unseren Algorithmus SCHNITTPUNKTSUCHE ohne Änderung einsetzen. Die Kenntnis der Punkte, in denen sich die Flächengrenzen schneiden, reicht aber offensichtlich nicht aus, um die planare Subdivision  $Sub(S_1, S_2)$  vollständig zu beschreiben. Erinnern wir uns zunächst an einige Definitionen aus Kap. 3.2. Eine planare Subdivision der Fläche wird u.a. anhand des Graphen

beschrieben, der dieser Subdivision zugrundeliegt. Dieser Graph ist durch die Flächengrenzen gegeben, die wiederum aus geraden Liniensegmenten (den Kanten des Graphs) und aus Ecken bestehen, an denen zwei Kanten aneinanderstoßen (den Knoten des Graphs). Das Ergebnis der Verschneidung von  $S_1$  und  $S_2$  ist die planare Subdivision  $Sub(S_1, S_2)$ , der wiederum ein Graph zugrundeliegt. Im Moment sind wir in der Lage, die Knoten dieses neuen Graphen effektiv zu bestimmen, nämlich als  $\{\text{Menge aller Knoten aus } S_1 \cup S_2\} \cup \{\text{Menge aller Schnittpunkte der Kanten aus } S_1 \cup S_2\}$ , und wir müssen uns im folgenden überlegen, wie zusätzlich auch die neuen Kanten, die neuen Teilflächen und die darauf bezogenen thematischen Daten berechnet werden können. Dafür benötigen wir zuallererst eine geeignete Repräsentation einer planaren Subdivision.

Bislang haben wir von einfachen geometrischen Strukturen (Punkten und Segmenten) gesprochen, ohne daß sich die Frage nach deren formaler Repräsentation zwingend gestellt hätte. Wenn wir dagegen Flächen betrachten, ändert sich die Lage ganz erheblich. Eine Fläche wurde beschrieben als eine Teilmenge der Ebene, die maximal verbunden ist und weder einen Punkt auf einer Kante noch einen Knoten beinhaltet. Wie kann ein solches Gebilde repräsentiert werden? Zur Beschreibung einer Fläche ist die Kenntnis aller sie umschließenden Knoten und Kanten notwendig, und – mehr noch – es muß die Reihenfolge bekannt sein, in der Kante auf Knoten auf Kante usw. folgt; mit anderen Worten, eine Repräsentation von Knoten und Kanten muß so gewählt sein, daß sie das ‘Umlaufen’ von Flächen ermöglicht. Ebenso muß sie z.B. gewährleisten, daß für einen gegebenen Knoten alle dort ‘eintreffenden’ Kanten oder für eine Fläche alle direkt benachbarten Flächen gefunden werden können, ohne aufwendig danach suchen zu müssen (auch wenn diese Anforderungen im Moment noch nicht unmittelbar nachvollziehbar sein mögen). Die Repräsentation, die im folgenden vorgestellt werden soll, unterstützt diese und andere notwendige Operationen und wird als *doppelt-verkettete Kantenliste* bezeichnet.

Die doppelt-verkettete Kantenliste besteht aus mindestens drei verschiedenen Tabellen, je eine für die Knoten, die Kanten und die Flächen einer planaren Subdivision (und kann beinahe beliebig erweitert werden; vgl. unten). Jeder Knoten, jede Kante und jede Fläche ist durch einen sog. *Eintrag* in der entsprechenden Tabelle repräsentiert. Eine Kante ist gewöhnlich an zwei Flächen gebunden, weshalb es naheliegend ist, jede Kante in Form von zwei *Halbkanten* zu speichern. Um das Umlaufen von Flächen gegen den Uhrzeigersinn zu ermöglichen, enthält jeder Eintrag der Kanten-Tabelle die nächst folgende Halbkante (wir müßten eigentlich von ‘Halbkanten-Tabelle’ sprechen). Um eine Fläche auch rückwärts, d.h. im Uhrzeigersinn umlaufen zu können, speichern wir ebenfalls die vorherige Halbkante. Die beiden Halbkanten, die zusammen eine Kante repräsentieren, nennen wir *Zwillinge*. Da Halbkanten eine Orientierung haben, sprechen wir vom *Ursprung* und vom *Ziel* einer Halbkante. Mit jeder Halbkante werden sowohl der *Zwilling* als auch der *Ursprung*(knoten) eingetragen (das Ziel der Halbkante ist dann gleich dem Ursprung des Zwillinges). Indem wir das vorwärtige Umlaufen einer Fläche als einen Umlauf gegen den Uhrzeigersinn festgelegt haben (was im Prinzip beliebig ist, aber ein für alle mal eindeutig sein muß), liegt die Fläche, die von dem Halbkantenzug umschlossen wird, aus Sicht eines Beobachters, der die Halbkanten in Laufrichtung abläuft, immer links. Die auf diese Weise links an eine Halbkante gebundene Fläche gehört ebenfalls zum Halbkanten-Eintrag.

Der Eintrag für einen Knoten enthält die beiden Koordinaten des Knotens sowie eine der Kanten, die mit dem Knoten verbunden ist. Der Eintrag für eine Fläche besteht schließlich aus einer beliebigen Halbkante, die die Fläche umschließt. Diese Halbkante nennen wir *äußere Halbkante*, und indem wir von dieser Kante ausgehend immer zur nächsten Kante gehen, können wir den äußeren Halbkantenzug der Fläche ablaufen und so die Fläche umrunden (die Fläche ist vollständig umrundet, wenn wir wieder an der äußeren Halbkante ankommen). Für Flächen mit Löchern reicht die Kenntnis der äußeren Halbkante jedoch nicht aus, vielmehr muß zusätzlich pro Loch je eine *innere Halbkante* für die Fläche eingetragen werden. Jeder innere Halbkantenzug, bestehend aus der inneren Halbkante und ihren nach-



Halbkante	Ursprung	Zwilling	geb.Fläche	Vorgänger	Nachfolger
$h_{11}$	$v_4$	$h_{12}$	$f_2$	$h_{31}$	$h_{21}$
$h_{12}$	$v_2$	$h_{11}$	$f_1$	$h_{52}$	$h_{41}$
$h_{21}$	$v_2$	$h_{22}$	$f_2$	$h_{11}$	$h_{31}$
$h_{22}$	$v_1$	$h_{21}$	$f_0$	$h_{32}$	$h_{51}$
$h_{31}$	$v_1$	$h_{32}$	$f_2$	$h_{21}$	$h_{11}$
$h_{32}$	$v_4$	$h_{31}$	$f_0$	$h_{42}$	$h_{22}$
$h_{41}$	$v_4$	$h_{42}$	$f_1$	$h_{12}$	$h_{52}$
$h_{42}$	$v_3$	$h_{41}$	$f_0$	$h_{51}$	$h_{32}$
$h_{51}$	$v_2$	$h_{52}$	$f_0$	$h_{22}$	$h_{42}$
$h_{52}$	$v_3$	$h_{51}$	$f_1$	$h_{41}$	$h_{12}$

Knoten	Koordinaten	geb.Kante
$v_1$	(1.0, 1.0)	$h_{31}$
$v_2$	(2.5, 6.0)	$h_{21}$
$v_3$	(6.5, 6.0)	$h_{52}$
$v_4$	(6.0, 1.0)	$h_{11}$

Fläche	äußere Begrenzung	innere Begrenzung
$f_0$	-	$h_{22}$
$f_1$	$h_{41}$	-
$f_2$	$h_{21}$	-

**Abbildung 3.7** Eine doppelt-verkettete Kantenliste. Eine doppelt-verkettete Kantenliste besteht aus drei Tabellen, mit denen die Knoten-, Kanten- und Flächeninformationen so verknüpft sind, daß alle wichtigen topologischen Relationen leicht zu ermitteln sind.

folgenden Halbkanten, umschließt das Loch und gehört somit zur Begrenzung der Fläche. Umgekehrt ist die Fläche an jede Halbkante dieses inneren Halbkantenzugs gebunden. Um die Konvention aufrechtzuerhalten, daß gebundene Flächen immer links von einer Halbkante liegen, muß sich der innere Halbkantenzug – und im Gegensatz zum äußeren Halbkantenzug! – *im Uhrzeigersinn* drehen.

Eine doppelt-verkettete Kantenliste speichert also neben den Koordinaten der Knoten überwiegend topologische Informationen: Welche Kante hat welchen Ursprung, welche Kante folgt auf eine andere oder geht ihr voraus, welche Fläche liegt links von einer Kante, welche Kanten gehören zur Begrenzung einer Fläche etc. Um eindeutige Bezüge zwischen den einzelnen Einträgen zu gewährleisten, enthält jeder Eintrag einen Identifikationsschlüssel, der sich von allen anderen Identifikationsschlüsseln innerhalb einer Tabelle unterscheidet. Die einzelnen Einträge können dann wie folgt beschrieben werden (vgl. Abb. 3.7):

- Jeder Eintrag für einen Knoten  $v$  enthält die Koordinaten  $v.x$  und  $v.y$  sowie einen Zeiger<sup>7</sup> *gebundeneKante(v)* auf eine beliebige Kante, die  $v$  als Ursprung hat.
- Jeder Eintrag für eine Halbkante  $h$  enthält einen Zeiger *Ursprung(h)* zu ihrem Ursprung, einen Zeiger *Zwilling(h)* zu ihrem Zwilling und einen Zeiger *gebundeneFläche(h)* zu der Fläche, von

<sup>7</sup>Der Begriff des ‘Zeigers’ wird hier ganz allgemein verwendet. Ein Zeiger in unserem Sinne kann sowohl auf eine bestimmte Speicheradresse im Arbeitsspeicher eines Rechners verweisen als auch, wie in diesem Fall, auf einen Tabelleneintrag anhand einer Identifikationsnummer. Die Tabelle wiederum, oder ein Teil von ihr, mag in einen Arbeitsspeicher geladen worden sein, und wir könnten mit einem Zeiger auf die Speicheradresse einer Identifikationsnummer verweisen oder gar auf die Adresse eines solchen Zeigers. Differenzierungen dieser Art sind für unsere Zwecke völlig überflüssig, so daß ohne Einschränkung auf die Einführung einer im informationstheoretischen Sinne korrekten Terminologie verzichtet werden kann.

deren Begrenzung sie ein Teil ist und die vom Ursprung aus gesehen links von ihr liegt. Es ist nicht notwendig, das Ziel einer Halbkante zu speichern, da dieses als  $Ursprung(Zwilling(h))$  leicht zu ermitteln ist (wenn zwei Halbkanten  $h_1$  und  $h_2$  Zwillinge sind, dann gilt immer sowohl  $h_1 = Zwilling(h_2)$  als auch  $h_2 = Zwilling(h_1)$ !). Zusätzlich werden die Zeiger  $Vorgänger(h)$  und  $Nachfolger(h)$  eingetragen, die die vorhergehende und nächste Kante der Kantenfolge bezeichnen, die die gebundene Fläche umschließt.  $Nachfolger(h)$  ist also die Kante auf der Begrenzung von  $gebundeneFläche(h)$ , die das Ziel von  $h$  als Ursprung hat, und  $Vorgänger(h)$  ist die Kante auf der Begrenzung von  $gebundeneFläche(h)$ , die  $Ursprung(h)$  als Ziel hat.

- Jeder Eintrag für eine Fläche  $f$  enthält einen Zeiger  $äußereHalbkante(f)$  zu einer beliebigen Halbkante, die auf der äußeren Begrenzung der Fläche liegt. Ist  $f$  eine ungebundene Fläche, setzen wir den Wert dieses Zeigers als 0. Zusätzlich wird eine Liste  $innereHalbkanten(f)$  geführt, die für jedes Loch in  $f$  genau einen Zeiger auf eine beliebige Halbkante enthält, die Teil der Begrenzung des Lochs ist. Die äußere Halbkante wird immer so gewählt, daß  $f$  links von ihr liegt, alle inneren Halbkanten so, daß  $f$  rechts von ihnen liegt.

Die Informationen in dieser doppelt-verketteten Kantenliste sind ausreichend für die grundlegenden Operationen, die wir im folgenden benötigen: Wir können eine Fläche umrunden, indem wir mit der Halbkante  $äußereHalbkante(f)$  beginnen und den  $Vorgänger(h)$ -Zeigern folgen; wir können alle Kanten, die an einen Knoten gebunden sind, im oder gegen den Uhrzeigersinn ‘besuchen’ oder alle Flächen, die zu einer Fläche direkt benachbart sind, ermitteln.

Die beschriebene doppelt-verkettete Kantenliste kann weiter vereinfacht werden, indem auf die Knoten-Tabelle verzichtet wird und die Knoten-Koordinaten direkt in die Halbkanten-Tabelle geschrieben werden (und zwar als Koordinaten des Ursprungs; man beachte, daß die Halbkanten-Tabelle über die Ursprünge mit der Knoten-Tabelle in einer 1:1-Beziehung steht). Für Anwendungen, bei denen Flächeninformationen keine Rolle spielen, kann auch auf die Flächen-Tabelle verzichtet werden. Häufig wird die doppelt-verkettete Kantenliste auch um Tabellen mit Zusatzinformationen erweitert (dabei handelt es sich z.B. um graphische Eigenschaften wie Farbe, Strichdicke, Strichlierung, Füllmuster usf. oder um thematische Daten). Dreh- und Angelpunkt jeder doppelt-verketteten Kantenliste bleibt jedoch immer die Tabelle der Halbkanten (ob mit oder ohne Koordinaten).

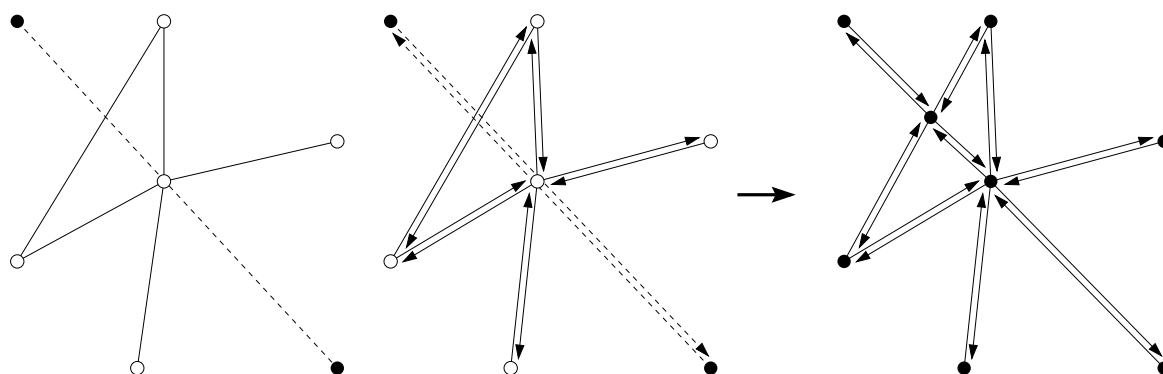
### 3.2.3 Berechnung der Schnittflächen aus zwei planaren Subdivisionen

Nachdem wir nun eine gute Repräsentation planarer Subdivisionen entworfen haben, können wir das allgemeine Kartenverschneidungsproblem endgültig in Angriff nehmen (vgl. S. 54). Die Aufgabe ist nun, für zwei gegebene doppelt-verkettete Kantenlisten der planaren Subdivisionen  $\mathcal{S}_1$  und  $\mathcal{S}_2$  die doppelt-verkettete Kantenliste für die planare Subdivision  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  zu bestimmen. Wir fordern zusätzlich, daß für jede Fläche  $f$  in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  die beiden Flächen aus  $\mathcal{S}_1$  und  $\mathcal{S}_2$  bekannt sind, die  $f$  enthalten. Auf diese Weise haben wir für jede ‘Schnittfläche’ Zugriff auf die thematischen Informationen der ‘Originalflächen’.<sup>8</sup>

Fragen wir uns zunächst, welche Informationen wir aus den vorhandenen doppelt-verketteten Kantenlisten unverändert übernehmen können. Da sind zunächst alle Knoten aus  $\mathcal{S}_1 \cap \mathcal{S}_2$ : Jeder ‘Originalknoten’ ist auch ein Knoten in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$ . Hinzu kommen alle Kanten aus  $\mathcal{S}_1 \cap \mathcal{S}_2$ , die sich mit keiner anderen Kante aus  $\mathcal{S}_1 \cap \mathcal{S}_2$  schneiden: Jede ‘Originalflächengrenze’ ist unverändert auch eine Flächengrenze in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$ . Die Knoten, die neu zu berechnen sind, sind die Schnittpunkte der Kanten aus

<sup>8</sup>Beweise für die Gültigkeit der hier vorgestellten Verfahren werden, falls notwendig, für die allgemeineren Definitionen in Kap. 3.3 an entsprechender Stelle nachgereicht.





**Abbildung 3.8** Die Aktualisierung der doppelt-verketteten Kantenliste  $\mathcal{D}$ . Zwei planare Subdivisionen (links) werden durch zwei doppelt-verkettete Kantenlisten repräsentiert, die nach  $\mathcal{D}$  kopiert werden. Die doppelt-verkettete Kantenliste  $\mathcal{D}$  ist zunächst ungültig, da sie keinen planaren Graph repräsentiert (Mitte). Aufgabe des Verschneidealgorithmus ist es,  $\mathcal{D}$  in eine gültige doppelt-verkettete Kantenliste zu transformieren (rechts).

$\mathcal{S}_1 \cap \mathcal{S}_2$ ; analog sind diejenigen Kanten neu zu berechnen, die sich dadurch ergeben, daß sich zwei Kanten aus  $\mathcal{S}_1 \cap \mathcal{S}_2$  schneiden. Wir kopieren daher zunächst die doppelt-verketteten Kantenlisten von  $\mathcal{S}_1$  und  $\mathcal{S}_2$  zusammen in eine neue doppelt-verkettete Kantenliste. Diese neue doppelt-verkettete Kantenliste ist natürlich zunächst keine gültige doppelt-verkettete Kantenliste, denn der ihr zugrundeliegende Graph ist nicht planar, da sich jetzt Kanten (aller Voraussicht nach) paarweise überschneiden. Die Aufgabe des gesuchten Verschneidealgorithmus ist es nun, diese ungültige doppelt-verkettete Kantenliste in eine gültige doppelt-verkettete Kantenliste zu transformieren, indem die Schnittpunkte gesucht und alle Knoten, Halbkanten und Flächen topologisch korrekt zusammengesetzt werden.

### 3.2.3.1 Aktualisierung der Knoten- und Kantentabelle

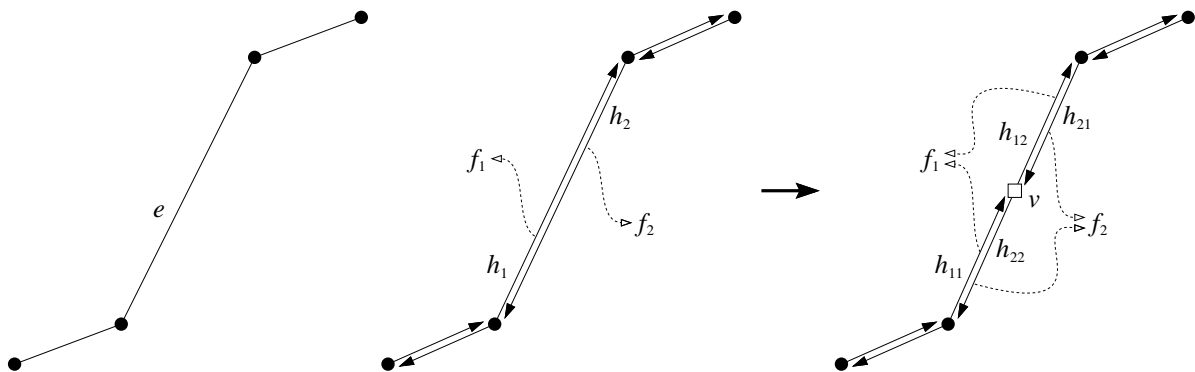
Betrachten wir zunächst die Frage, wie die Knoten- und Kantentabelle der doppelt-verketteten Kantenliste von  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  bestimmt werden; die Aktualisierung der Flächeneinträge heben wir uns für später auf. Die Grundlage für unseren Algorithmus ist natürlich das Scanline-Verfahren; wie im Kap. 3.2.1 erläutert, basiert dieses Verfahren auf einer Eventfolge  $\mathcal{Q}$ , die die einzelnen Eventpunkte speichert (in der Reihenfolge von oben nach unten und im Zweifelsfall von links nach rechts), sowie auf einer Statusfolge  $\mathcal{T}$ , die alle Segmente enthält, die an einem bestimmten Eventpunkt die Scanlinie schneiden (in der Reihenfolge von links nach rechts und im Zweifelsfall kurz unterhalb der Scanlinie). Nun stehen uns, da wir mit doppelt-verketteten Kantenlisten arbeiten, keine Segmente im eigentlichen Sinne zur Verfügung, sondern Halbkanten, die sich paarweise zur Beschreibung einer Kante ergänzen. Um den Scanline-Algorithmus anwenden zu können, müssen wir die Halbkantenpaare ( $h$  und  $Zwilling(h)$ ) zu je einer Kante  $e$  zusammenfassen. Diese Kanten betrachten wir, genau wie wir das oben bei den Segmenten bereits implizit getan haben, als geschlossen in dem Sinne, daß die Knoten mit zur Kante gehören. Mit anderen Worten: Unter einer Kante  $e = (h_1, h_2)$ ,  $h_1 = Zwilling(h_2)$  verstehen wir im folgenden das Liniensegment (die Strecke)  $\overline{Ursprung(h_1)Ursprung(h_2)}$ .<sup>9</sup>

<sup>9</sup>Strenggenommen ist diese Terminologie natürlich nicht korrekt, da mit dem Begriff der ‘Kante’ i.d.R. eine zweistellige Relation über der Menge der Knoten eines Graphs gemeint ist, und nicht – wie hier – die Menge aller Punkte, die zwei Knoten geradlinig miteinander verbinden, inklusive der Knoten(punkte) selbst. Um jedoch die Beziehung zwischen einer planaren Subdivision, dem zugrundeliegenden planaren Graphen und einer doppelt-verketteten Kantenliste nicht aus den Augen zu verlieren, soll im folgenden dennoch, und analog zu de Berg et al. (1997), von Kanten und nicht von Liniensegmenten gesprochen werden. In ähnlich umgangssprachlicher Weise verwenden wir auch die Begriffe ‘Punkt’ und ‘Knoten’ synonym.

Nachdem wir die doppelt-verketteten Kantenlisten von  $\mathcal{S}_1$  und  $\mathcal{S}_2$  kopiert haben, erhalten wir eine neue doppelt-verkettete Kantenliste, die wir mit  $\mathcal{D}$  bezeichnen wollen. Bearbeiten wir nun alle Kanten aus  $\mathcal{D}$  mit unserem Algorithmus SCHNITTPUNKTSUCHE, dann finden wir sämtliche Knoten, in denen sich die Kanten schneiden. Zusammen mit den Knoten, die bereits in  $\mathcal{D}$  vorhanden sind, sind das alle Knoten, die in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  existieren und demnach nach Abschluß des Verfahrens in  $\mathcal{D}$  vorhanden sein müssen. Wenn wir nun jeden Schnittpunkt  $v$ , anstatt ihn einfach nur zu berichten, zusammen mit einer beliebigen an  $v$  gebundenen Kante in  $\mathcal{D}$  einfügen (vgl. Zeile 4 in BEARBEITEEVENTPUNKT), haben wir die Knotentabelle von  $\mathcal{D}$  bereits vollständig aktualisiert. Die Aktualisierung der Kantentabelle von  $\mathcal{D}$  kann ebenso bei der Bearbeitung eines Eventpunktes erfolgen, erfordert aber etwas mehr Aufwand. Stellen wir uns zwei Kanten vor, die sich in einem Punkt schneiden, der nicht ein Endpunkt dieser Kanten ist. Wir sind nun nicht nur an dem Schnittpunkt selbst interessiert, sondern fordern, daß sich die beiden Kanten gegenseitig wirklich ‘zerschneiden’ (Abb. 3.8). Aus den zwei Kanten müssen also vier entstehen, die im Nachhinein alle an denselben Knoten (den Schnittpunkt) gebunden sind. Etwas allgemeiner ausgedrückt heißt das, daß jede Kante, die einen Schnittpunkt  $v$  enthält, der nicht ein Endpunkt dieser Kante ist, in  $v$  *geteilt* werden muß (man beachte, daß die Teilung einer Kante in  $v$  ausgeschlossen ist, wenn  $v$  ein Endpunkt dieser Kante ist). Da eine Kante aus zwei Halbkanten besteht, entstehen beim Teilen einer Kante vier neue Halbkanten, und es müssen zwei Einträge in der Kantentabelle von  $\mathcal{D}$  durch vier neue Einträge ersetzt werden.

de Berg et al. (1997, S. 35) schlagen vor, die beiden zu teilenden Halbkanten weiter zu benutzen und zusätzlich nur zwei neue zu erzeugen. Dieses Vorgehen ist für uns mit Hinblick auf die Überlegungen in Kap. 3.5 jedoch mit großen Schwierigkeiten verbunden, weshalb wir die Kanten und die entsprechenden Einträge in  $\mathcal{D}$  grundsätzlich niemals ändern, sondern immer komplett ersetzen. Da wir die Eventfolge  $\mathcal{Q}$  später ein zweites mal durchlaufen wollen, müssen wir daher auch einen Teil der Eventpunkte aktualisieren: Mit jedem Eventpunkt  $v$  sind ja die Kanten gespeichert, deren oberer Endpunkt in  $v$  liegt. Teilen wir nun eine Kante an einem Schnittpunkt  $v_i$  unterhalb von  $v$  ( $v \prec_{\mathcal{Q}} v_i$ ), so müssen wir die zu teilende Kante (die mit  $v$  gespeichert ist) von  $v$  entfernen und  $v$  die neue (Teil-)Kante  $(v, v_i)$  zuweisen. Die Suche in  $\mathcal{Q}$  nach einem Eventpunkt  $v$  benötigt  $\log(n_v + I)$  Schritte, wobei  $n_v$  die Anzahl der Knoten und  $I$  die Anzahl der Schnittpunkte in  $\mathcal{S}_1 \cup \mathcal{S}_2$  ist. Die Anzahl der notwendigen Aktualisierungen von  $\mathcal{Q}$  ist linear zum Grad des Schnittpunktes  $v_i$ . Diese Vorgehensweise hat auf die Laufzeit und den Speicherbedarf unseres Algorithmus keinen Einfluß (vgl. Kap. 3.3).

Was muß nun bei der Teilung einer Kante  $e = (h_1, h_2)$  in  $v$  geschehen? Zunächst werden für jede der beiden Halbkanten  $h_i$  ( $i = 1, 2$ ), zwei neue Halbkanten  $h_{i1}$  und  $h_{i2}$  erzeugt, so daß  $Ursprung(h_{i1}) = Ursprung(h_i)$  und  $Ursprung(h_{i2}) = v$ . Die Flächeninformationen werden kopiert, so daß  $gebundeneFläche(h_{ij}) = gebundeneFläche(h_i)$  ( $j = 1, 2$ ), und die Zwillinge werden zugewiesen, so daß  $h_{11} = Zwillig(h_{22})$  und  $h_{12} = Zwillig(h_{21})$  (Abb. 3.9). Diese Beziehungen gelten für alle Kanten, die in  $v$  geteilt werden. Bei der Aktualisierung der *Vorgänger*()- und *Nachfolger*()-Zeiger muß man dagegen unterscheiden zwischen der Situation um die Endpunkte von  $e$  und der Situation um den Schnittpunkt  $v$ . Die Aktualisierung in den Endpunkten von  $e$  ist einfach, denn der Vorgänger der Halbkante  $h_i$  wird zum Vorgänger der Halbkante  $h_{i1}$  und anschließend die Halbkante  $h_{i1}$  zum Nachfolger ihres Vorgängers erklärt. Die Aktualisierung der Kanten um  $v$  ist dagegen etwas komplizierter. Wir haben gesehen, daß die Teilung einer Kante  $e_1$  in  $v$  dann und nur dann notwendig wird, wenn (1.)  $v$  nicht ein Endpunkt von  $e_1$  ist und (2.) es mindestens eine Kante  $e_2 \neq e_1$  gibt, die  $e_1$  in  $v$  schneidet bzw. ‘berührt’ (die in Abb. 3.9 dargestellte Situation ist dementsprechend unvollständig, denn es fehlt mindestens eine zweite Kante, durch die die Existenz von  $v$  erst möglich wird). Erinnern wir uns auch, daß ein Grund für die Einführung von ‘Halbkanten’ der war, daß Flächen umlaufen werden können. Die *Vorgänger*()- und *Nachfolger*()-Zeiger in einem Schnittpunkt  $v$  müssen daher auch so korrigiert werden, daß die (Schnitt-)flächen in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  korrekt umlaufen werden können. Wenn wir schließlich noch unsere

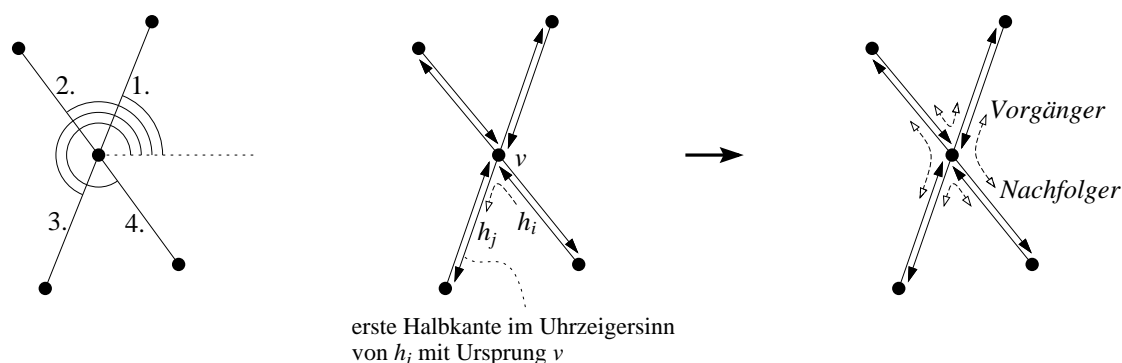


**Abbildung 3.9** Die Teilung einer Kante. Ein Kantenzug (links) wird durch eine doppelt-verkettete Kantenliste repräsentiert (Mitte). Wird eine Kante an einem Schnittpunkt geteilt, müssen die Vorgänger- und Nachfolger-Beziehungen aktualisiert und die Flächeninformationen kopiert werden (rechts). Man beachte, daß die dargestellte Situation in dem Sinne unvollständig ist, als daß mindestens eine zweite Kante fehlt, durch die die Existenz des Schnittpunkts  $v$  erst möglich wird.

Konvention bzgl. der Orientierung von Halbkantenzügen berücksichtigen, nach der die Außengrenzen von Flächen immer gegen den Uhrzeigersinn orientiert sind, ist die Situation eindeutig: wenn wir einer Halbkante und ihren Nachfolgern folgen und an einen Schnittpunkt gelangen, müssen wir immer ‘die erste links’ abbiegen (die Gültigkeit dieses einfachen Prinzips ist bereits in Abb. 3.8 gut zu sehen).

An einem Eventpunkt  $v$  teilen wir also zunächst alle Kanten, die  $v$  enthalten und deren Endpunkte beide von  $v$  verschieden sind. Alle am Eventpunkt  $v$  involvierten Kanten sind dann an  $v$  gebunden, und wir können anhand des Winkels, den sie mit der Horizontalen einschließen, eine eindeutige Ordnung angeben, in der die Kanten um  $v$  herum liegen (vgl. aber auch Kap. 3.5.1.2). Für die Hälfte der entsprechenden Halbkanten liegt der Ursprung in  $v$ , die andere Hälfte hat  $v$  zum Ziel. Folgen wir nun dem ‘Linksabbieger-Prinzip’, dann muß für jede Kante  $h_i$ , die  $v$  zum Ziel hat, diejenige Kante  $h_j$  der *Nachfolger*( $h_i$ ) werden, deren Ursprung  $v$  ist und die von  $h_i$  aus gesehen im Uhrzeigersinn benachbart ist. Entsprechend ist auch *Vorgänger*( $h_j$ ) :=  $h_i$  (Abb. 3.10).

Mit Ausnahme dieses letzten Schritts benötigen alle zur Teilung von Kanten und zur Aktualisierung der Einträge in  $\mathcal{D}$  notwendigen Schritte konstante Zeit. Nur die Korrektur der Situation um einen Schnittpunkt  $v$  dauert länger, und zwar in Abhängigkeit des Grades von  $v$ . Wir haben mit Hinblick auf die Diskussion in Kap. 3.3 an dieser Stelle bereits die allgemeine Lösung angegeben, die die Kenntnis der zyklischen Ordnung aller Kanten um  $v$  voraussetzt. Da diese Ordnung eindeutig ist, können wir einen balancierten Binärbaum verwenden, dessen Initialisierung  $O(n_v \log n_v)$  Schritte erfordert und in dem in  $O(\log n_v)$  nach einem Nachbarn gesucht werden kann, wobei  $n_v$  der Grad von  $v$  ist. Die Laufzeit des Verschneide-Algorithmus steigt dadurch asymptotisch, soll aber an dieser Stelle noch nicht weiter untersucht werden. Bei der Verschneidung von nur zwei planaren Subdivisionen  $\mathcal{S}_1$  und  $\mathcal{S}_2$  ist die Kenntnis der zyklischen Ordnung aller Kanten um  $v$  nicht notwendig, da man sich die ‘Zugehörigkeit’ der Kanten (die ja entweder zu  $\mathcal{S}_1$  oder  $\mathcal{S}_2$  gehören) zunutze machen kann: Es genügt, die benachbarten Paare der sich in  $v$  schneidenden Kanten zu bestimmen, so daß eine Kante zu  $\mathcal{S}_1$  und die andere zu  $\mathcal{S}_2$  gehört. Die Korrektur der Situation um einen Schnittpunkt  $v$  benötigt dann eine Zeit, die linear zum Grad von  $v$  ist, und die für die Verschneidung von zwei planaren Subdivisionen benötigte Zeit liegt in  $O(n \log n + k \log k)$ , wobei  $n$  die Summe der Komplexitäten von  $\mathcal{S}_1$  und  $\mathcal{S}_2$  und  $k$  die Komplexität von  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  ist (vgl. de Berg et al. 1997, S. 35f.).



**Abbildung 3.10** Das ‘Linksabbieger-Prinzip’ beim Aktualisieren der Flächeninformation. Alle Kanten, die an einen Knoten gebunden sind, können anhand des Winkels, den sie mit der Horizontalen einschließen, in eine eindeutige Ordnung gebracht werden (links). Eine beliebige Kante begrenzt dieselbe Fläche wie ihre im Uhrzeigersinn benachbarte Kante. Dementsprechend müssen die *Vorgänger*(-) und *Nachfolger*(-)Zeiger gesetzt werden (Mitte und rechts).

### 3.2.3.2 Aktualisierung der Flächentabelle

Nachdem alle Knoten und Kanten von  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  bekannt sind, können wir uns der Berechnung der Flächeninformationen für  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  zuwenden. Konkret bedeutet das, für jede Fläche  $f$  in  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  genau einen Eintrag in der Flächentabelle von  $\mathcal{D}$  vorzunehmen, so daß *äußereBegrenzung*( $f$ ) auf eine beliebige Halbkante zeigt, die auf der äußeren Grenze von  $f$  liegt, und *innereBegrenzung*( $f$ ) eine Liste von Zeigern enthält, die auf je eine beliebige Halbkante auf der äußeren Grenze jedes Lochs in  $f$  verweisen. Zusätzlich muß für jede Halbkante  $h$ , die auf der äußeren oder inneren Grenze von  $f$  liegt, der Eintrag *innereFläche*( $h$ ) auf  $f$  zeigen. Wir nehmen diese Einträge erst dann vor, wenn der  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  zugrundeliegende Graph korrekt berechnet ist, wenn also die Eventfolge  $\mathcal{Q}$  vollständig abgearbeitet ist und alle bis hierher besprochenen Einträge in die Knoten- und Kantentabelle von  $\mathcal{D}$  erfolgt sind. Es ist klar, daß jede Begrenzung von  $f$  (sowohl die äußere als auch alle inneren) geschlossen sein muß: Wenn wir von einer beliebigen Kante  $h$  ausgehen und, den *Nachfolger*(-)Zeigern folgend, solange immer zur nächsten Kante gehen, bis wir wieder in  $h$  ankommen, können wir jeden  $f$  begrenzenden Halbkantenzug ablaufen und  $f$  (gegen den Uhrzeigersinn) bzw. ein Loch von  $f$  (im Uhrzeigersinn) umrunden.<sup>10</sup> Um die einzelnen Kantenzüge zu finden, reicht es aus, den planaren Graphen von  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  einmal zu durchqueren. Zusätzlich ist jedoch zu klären, welche dieser einzelnen Kantenzüge eine äußere Begrenzung darstellen, welche zu einer inneren Begrenzung gehören, d.h. ein Loch umgrenzen, und welche Löcher innerhalb welcher äußeren Grenzen liegen.

Um zunächst zu entscheiden, ob ein Halbkantenzug eine äußere oder eine innere Begrenzung ist, müssen wir seine Orientierung ermitteln. Dazu betrachten wir jeweils den am weitesten links stehenden Knoten  $v_l$  des Halbkantenzuges (gibt es mehrere solcher Knoten, wählen wir den obersten). Alle Halbkanten sind so ausgerichtet, daß die von ihnen eingeschlossene (die gebundene) Fläche links von ihnen liegt. Es gibt genau eine Halbkante, die  $v_l$  zum Ziel hat, und genau eine, deren Ursprung in  $v_l$  liegt. Ist nun der Winkel in  $v_l$ , den diese beiden Halbkanten links von sich einschließen, kleiner als  $180^\circ$ , verläuft der Halbkantenzug gegen den Uhrzeigersinn und ist damit eine äußere Begrenzung; ist

<sup>10</sup>Sollten wir bei der Berechnung von  $\mathcal{D}$  einen Fehler gemacht haben, würde dies fast unweigerlich dazu führen, daß wir uns in einigen Fällen in einem Kantenzug verlaufen, der  $f$  nicht begrenzt. In einem solchen Fall würden wir nie mehr zu  $h$  zurückfinden; das Verfahren würde dann nie terminieren und unser Algorithmus wäre wertlos (vgl. Kap. 3.5).

der Winkel dagegen größer als  $180^\circ$ , ist der Halbkantenzug im Uhrzeigersinn orientiert, er umgrenzt also ein Loch. Diese Eigenschaft gilt für den am weitesten links stehenden Knoten eines jeden Halbkantenzuges, aber nicht notwendigerweise für jeden Knoten (Abb. 3.11 links).<sup>11</sup>

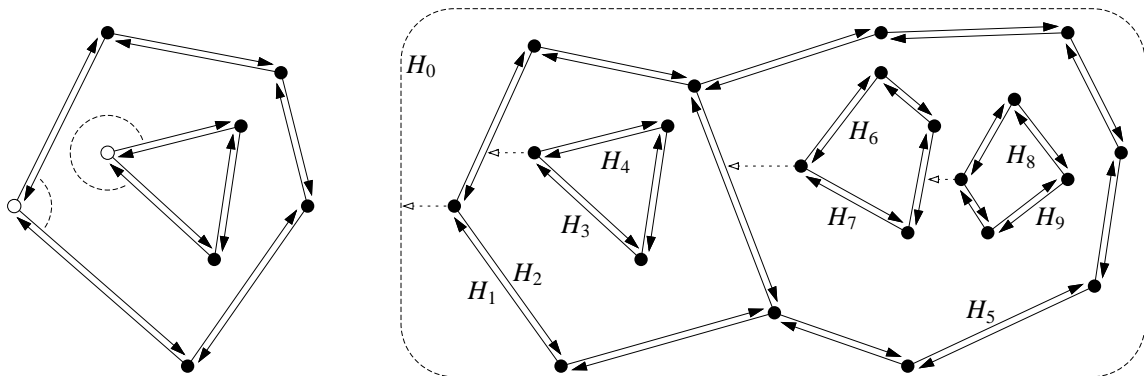
Die Frage, welche Halbkantenzüge dieselbe Fläche umgrenzen, beantworten wir mit Blick auf die Diskussion in Kap. 3.3 abweichend von de Berg et al. (1997, S. 36ff.) wie folgt: Angenommen, uns steht die Eventfolge  $\mathcal{Q}$  noch zur Verfügung, nachdem sie bereits einmal abgearbeitet wurde (und das tut sie aufgrund unserer Definition des Algorithmus SCHNITTPUNKTSUCHE tatsächlich; vgl. S. 59 und Fn. 6).  $\mathcal{Q}$  enthält dann alle Knoten aus  $\mathcal{D}$ , also incl. aller Schnittpunkte, als Eventpunkte entsprechend der Ordnung  $\prec_{\mathcal{Q}}$  (vgl. S. 58), und alle Halbkanten sind in  $\mathcal{D}$  korrekt eingetragen. Wir durchlaufen die Eventfolge  $\mathcal{Q}$  nun einfach ein zweites mal<sup>12</sup> und führen an jedem Eventpunkt  $v$  folgende Operationen durch: Sei  $U(v)$  die Menge der Halbkanten, deren Ursprung in  $v$  liegt und deren Ziel eine niedrigere Priorität als der Ursprung hat (für jede Halbkante  $h \in U(v)$  ist also  $\text{Ursprung}(h) \prec_{\mathcal{Q}} \text{Ziel}(h)$ , d.h.,  $v$  ist der ‘obere Endpunkt’ von  $h$ ). Es ist klar, daß jede Halbkante aus  $U(v)$  eine andere Fläche aus  $\text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$  begrenzt; die Anzahl der durch die Halbkanten aus  $U(v)$  begrenzten Flächen ist also gleich der Anzahl der Halbkanten in  $U(v)$ . Wir stellen nun für jede Halbkante  $h$  aus  $U(v)$ , deren *innereFläche()*-Zeiger noch nicht auf einen Flächeneintrag in  $\mathcal{D}$  verweist, die Orientierung des in  $h$  beginnenden Kantenzuges fest. Ist der Kantenzug gegen den Uhrzeigersinn orientiert, haben wir eine äußere Begrenzung gefunden. Wir erzeugen einen neuen Flächeneintrag, fügen ihn in  $\mathcal{D}$  ein und weisen die *innereFläche()*-Zeiger aller Halbkanten des Kantenzuges diesem Eintrag zu. Ist der Kantenzug dagegen im Uhrzeigersinn orientiert, haben wir ein Loch gefunden und müssen nun entscheiden, innerhalb welcher Fläche, d.h. innerhalb welcher äußeren Grenzen es liegt.

Das Entscheidungsverfahren verdeutlicht Abb. 3.11 (rechts): Sei  $v_l$  der am weitesten links liegende Knoten des Kantenzuges durch  $h$ ,  $e_l = (h_u, h_d)$  die Kante, die direkt links von  $v_l$  liegt, und  $h_d$  die Halbkante von  $e_l$ , deren Ursprung im oberen Endpunkt von  $e_l$  liegt. Betrachten wir die Orientierung des Kantenzuges, in dem  $h_d$  liegt, so können wir folgendes feststellen: Ist der Kantenzug durch  $h_d$  gegen den Uhrzeigersinn orientiert, haben wir die äußere Begrenzung der Fläche gefunden, innerhalb derer der Kantenzug durch  $h$  ein Loch bildet. Ist der Halbkantenzug durch  $h_d$  dagegen im Uhrzeigersinn orientiert, haben wir ein weiteres Loch gefunden und müssen die Suche nach der äußeren Begrenzung der Fläche, in der es liegt, fortsetzen. In jedem Fall liegt dieses zweite Loch, das durch  $h_d$  begrenzt wird, innerhalb derselben Fläche wie das Loch, das durch  $h$  begrenzt wird. Auf diese Weise ‘hangeln’ wir uns solange nach links von einem Loch zum nächsten, bis wir auf eine Halbkante  $h_d$  stoßen, die Teil eines gegen den Uhrzeigersinn orientierten Halbkantenzuges ist. In diesem Fall verweist der *innereFläche()*-Zeiger der Halbkante  $h_d$  auf die gesuchte Fläche, denn da wir der Ordnung der Eventfolge  $\mathcal{Q}$  folgen, finden wir aufgrund der Invariante des Scanline-Prinzips von jeder Fläche immer zuerst die äußere Begrenzung und erst dann deren Löcher. Wir fügen je eine beliebige Halbkante jedes gefundenen Lochs zur *innereBegrenzung()*-Liste der gesuchten Fläche in  $\mathcal{D}$  hinzu und weisen den *innereFläche()*-Zeigern aller Halbkanten auf den Begrenzungen der Löcher die Fläche zu.

Für den Fall, daß keine Halbkante links von einem Knoten liegt, ist die das Loch umschließende Fläche (nach der wir ja suchen) die ungebundene Fläche. Hier wird nun die Bedeutung der ungebun-

<sup>11</sup>Man beachte, daß dieser Winkel niemals genau gleich  $180^\circ$  sein kann und immer auch größer als  $0^\circ$  ist, vorausgesetzt, der Halbkantenzug enthält mindestens drei sich paarweise nicht überlappende Halbkanten (andernfalls würde ja keine Fläche begrenzt).

<sup>12</sup>Der Vorschlag von de Berg et al. (1997, S. 36ff.) ist dem unsrigen sehr ähnlich, nur wird bei ihnen die Eventfolge nur einmal durchlaufen. Mit ihrem Ansatz werden äußere und innere Grenzen in einer unsystematischen Reihenfolge gefunden, weshalb sie auf die Konstruktion eines Hilfsgraphen angewiesen sind. Wird die Eventfolge  $\mathcal{Q}$  ein zweites mal durchlaufen, wird immer erst die äußere Begrenzung einer Fläche gefunden und erst danach die Begrenzungen ihrer Löcher. Solange die Anzahl der Durchläufe von  $\mathcal{Q}$  konstant ist (hier: 2), hat ein mehrfaches Abarbeiten von  $\mathcal{Q}$  auf die Laufzeit des gesamten Verfahrens natürlich keinen Einfluß.



**Abbildung 3.11** Die Zuordnung von äußeren und inneren Flächengrenzen. Die Orientierung eines Kantenzuges kann anhand des Winkels, der in dem am weitesten links liegenden Knoten eingeschlossen wird, bestimmt werden (links). Die Zuordnung von äußeren Grenzen (gegen den Uhrzeigersinn orientiert) und Löchern (im Uhrzeigersinn orientiert) ist möglich, wenn für diese Knoten auch die jeweils links benachbarte Kante bekannt ist (rechts).

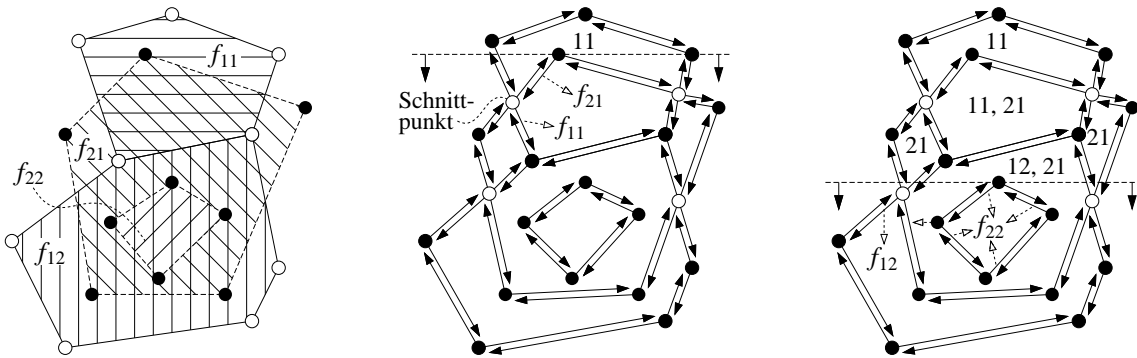
denen Fläche klar: Sie ist notwendig, um *allen* Halbkanten der doppelt-verketteten Kantenliste einer planaren Subdivision  $\mathcal{S}$  einen definierten Wert für den *innereFläche()*-Zeiger zuweisen zu können; sie beschreibt die Euklidische Ebene  $E$ , in der jede in  $E$  vollständig verbundene Fläche der Vereinigungsmenge aller mit  $\mathcal{S}$  gegebenen Flächen ein Loch darstellt; sie ist schließlich auch wertvoll, weil durch die Umkehr der Orientierung der Begrenzungen dieser Löcher leicht der gemeinsame Umriß aller Flächen aus  $\mathcal{S}$  konstruiert werden kann (die Lösung des Verschneideproblems liefert also die Lösung des ‘Vereinigungsproblems’ gleich mit).

Wie finden wir aber für jeden Knoten  $v$  die direkt links von  $v$  benachbarte, von oben nach unten verlaufende Halbkante  $h_d$ ? Die Antwort ist einfach, wenn wir das Problem präzisieren und die Frage nach der links benachbarten Kante eines aktuellen Eventpunktes  $v$  stellen. Diese Kante muß natürlich am Eventpunkt  $v$  in der Statusfolge  $\mathcal{T}$  enthalten sein. Wir finden sie also schon während des ersten ‘Scans’, d.h. beim ersten Durchlaufen der Eventfolge  $\mathcal{Q}$ : Sie entspricht dem Segment  $s_l$  in den Zeilen 10 bzw. 13 in BEARBEITEEVENTPUNKT. Wir finden die links von  $v$  benachbarte Kante also ohne zusätzliche Kosten und müssen sie uns nur mit  $v$  zusammen merken. Die für das Durchlaufen eines einzelnen, geschlossenen Kantenzuges benötigte Zeit ist linear zur Länge des Kantenzuges, und jeder Kantenzug wird genau zweimal durchlaufen (einmal, um die Orientierung festzustellen, ein zweites mal, um die *innereFläche()*-Zeiger zu setzen). Für jeden geschlossenen Kantenzug erfolgt genau ein Eintrag in  $\mathcal{D}$ , jeder Eintrag benötigt konstante Zeit. Das erneute Durchlaufen von  $\mathcal{Q}$  benötigt  $(n + I)$  Schritte, wobei  $I$  die Anzahl der Schnittpunkte in  $\mathcal{S}_1 \cup \mathcal{S}_2$  ist.<sup>13</sup> Die zur Aktualisierung der Flächeninformationen einer planaren Subdivision  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  der Komplexität  $k$  zusätzlich benötigte Zeit liegt daher in  $O(n + k)$ .

### 3.2.3.3 Aktualisierung der Flächenlabel

Was bleibt, ist die Aufgabe, jeder Fläche der Verschneidung die beiden Flächen der originalen Subdivisionen zuzuordnen, aus denen sie entstanden ist. Zu bestimmen sind also die beiden Flächen  $f_1 \in \mathcal{S}_1$  und  $f_2 \in \mathcal{S}_2$ , die die Fläche  $f \in Sub(\mathcal{S}_1, \mathcal{S}_2)$  enthalten. Der Lösungsweg beruht auf folgender Über-

<sup>13</sup>Da wir die Eventfolge  $\mathcal{Q}$  als balancierten Binärbaum implementiert haben, erfordert das vollständige Durchqueren von  $\mathcal{Q}$  eigentlich  $(n + I) \log(n + I)$  Schritte. Da  $\mathcal{Q}$  jedoch nicht mehr verändert werden muß, transformieren wir  $\mathcal{Q}$  in eine Liste. Die Transformation benötigt  $(n + I)$  Schritte, das anschließende Durchqueren dieser Liste ebenfalls  $(n + I)$  Schritte.



**Abbildung 3.12** Die Bestimmung der Flächenlabel. 'Ursprungsflächen' (links) werden während der Aktualisierung der Flächeninformation durch die *innereFläche()*-Zeiger der doppelt-verketteten Kantenliste  $\mathcal{D}$  repräsentiert. An einem Schnittpunkt können die Flächeninformationen direkt an den *innereFläche()*-Zeigern der entsprechenden Halbkanten abgelesen werden (Mitte). Gibt es keinen Schnittpunkt innerhalb eines Halbkantenzugs, muß die einen Knoten umschließende Fläche bekannt sein (rechts).

legung: Betrachte einen beliebigen Knoten  $v$  auf der Begrenzung einer Fläche  $f \in \text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$ . Wenn  $v$  ein Schnittpunkt ist, können die beiden Flächen aus  $\mathcal{S}_1$  und  $\mathcal{S}_2$  direkt an den *innereFläche()*-Zeigern der entsprechenden Halbkanten abgelesen werden. Da bei der Verschneidung von zwei planaren Subdivisionen maximal zwei Flächen 'übereinanderliegen' können, können die Label für jede Fläche  $f$ , deren Begrenzung mindestens einen Schnittpunkt enthält, auf diese Weise bestimmt werden (Abb. 3.12 Mitte; man beachte, daß auch Berührungspunkte zu den Schnittpunkte zählen!). Problematisch sind nun noch jene Flächen aus  $\mathcal{S}_1$  ( $\mathcal{S}_2$ ), die vollständig von einer Fläche aus  $\mathcal{S}_2$  ( $\mathcal{S}_1$ ) eingeschlossen werden (Abb. 3.12 rechts). In diesem Fall bildet die eingeschlossene Fläche ein Loch innerhalb der sie umschließenden Fläche, und wir können letztere an keinem Knoten  $v$  aus den *innereFläche()*-Zeigern ableiten (Abb. 3.12 rechts). Damit ist auch schon klar, was zu tun ist: Wir müssen bestimmen, in welcher Fläche aus  $\mathcal{S}_1$  ( $\mathcal{S}_2$ ) ein beliebiger Knoten  $v$  aus  $\mathcal{S}_2$  ( $\mathcal{S}_1$ ) liegt. Ist diese Fläche für jeden Knoten aus  $\text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$  bekannt, lassen sich für jede Fläche aus  $\text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$  die Label der beiden Originalflächen auf einfache Weise bestimmen. Die gesuchten einschließenden Flächen können wir für jeden Knoten während des ersten 'Scans' wie folgt ermitteln:<sup>14</sup> Wir kennen bereits für jeden Eventpunkt  $v$  die in der Statusfolge  $\mathcal{T}$  direkt links benachbarte Kante  $e_l = (h_u, h_d)$  und damit auch die Halbkante  $h_d$ , deren Ursprung im oberen Endpunkt von  $e_l$  liegt, so daß also  $\text{Ursprung}(h_d) \prec_Q \text{Ursprung}(h_u)$ . Nehmen wir an,  $v$  sei kein Schnittpunkt und gehöre ursprünglich zur planaren Subdivision  $\mathcal{S}_i, i = (1, 2)$ . Wenn nun  $h_d$  ursprünglich zur planaren Subdivision  $\mathcal{S}_j$  gehört ( $j = (1, 2)$  und  $j \neq i$ ), zeigt *innereFläche*( $h_d$ ) auf die gesuchte Fläche, die  $v$  einschließt. Gehört  $h_d$  dagegen ursprünglich zu  $\mathcal{S}_i$ , also zur selben planaren Subdivision wie  $v$ , dann gibt es zwei Möglichkeiten: Wenn  $\text{Ursprung}(h_d)$  kein Schnittpunkt ist, wird  $v$  von derselben Fläche umschlossen wie  $\text{Ursprung}(h_d)$ . Die  $\text{Ursprung}(h_d)$  umschließende Fläche ist wegen der Invariante des Scanline-Prinzips bekannt, denn  $\text{Ursprung}(h_d)$  liegt in der Eventfolge  $\mathcal{Q}$  vor  $v$ .<sup>15</sup> Ist  $\text{Ursprung}(h_d)$  dagegen ein Schnittpunkt, verläuft die Grenze der gesuchten Fläche durch  $\text{Ursprung}(h_d)$ : Sei  $U(h_d)$  die Menge der Kanten aus  $\mathcal{S}_j$ , die einen Endpunkt in  $\text{Ursprung}(h_d)$  haben. Ordnen wir die Kanten aus  $U(h_d)$  gegen den Uhrzeigersinn um  $\text{Ursprung}(h_d)$ , dann wird die gesuchte Fläche von derjenigen Kante  $e_{ll}$  begrenzt, die in dieser Ordnung im Uhrzeigersinn zu  $e_l$  benachbart ist. Entsprechend zeigt der *innereFläche()*-Zeiger derjenigen Halbkante von  $e_{ll}$ , deren Ursprung in

<sup>14</sup>de Berg et al. (1997, S. 38) verweisen an dieser Stelle nur auf das Scanline-Prinzip und überlassen die Lösung ihren Lesern. Zum Beweis des allgemeinen Falls vgl. S. 80ff.

<sup>15</sup> $h_d$  liegt per definitionem in  $\mathcal{T}$  links von  $v$ . Daher gilt immer auch  $\text{Ursprung}(h_d) \prec_Q v$ !

$Ursprung(h_d)$  liegt, auf die gesuchte Fläche (Abb. 3.12 rechts; vgl. auch Abb. 3.10 links). Die Suche nach der einschließenden Fläche für einen Eventpunkt  $v$  benötigt in den ersten beiden Fällen konstante Zeit, im letzten Fall eine Zeit, die linear zum Grad eines Schnittpunktes oberhalb von  $v$  ist. Die Gesamtlaufzeit des Verfahrens wird daher nicht beeinträchtigt.

### 3.2.3.4 Ein einfacher Verschneidealgorithmus

Nun haben wir endlich alle Informationen zusammen, die wir benötigen, um eine ungültige doppelt-verkettete Kantenliste  $\mathcal{D}$  (die Kopie zweier gültiger doppelt-verketteter Kantenlisten) in eine gültige doppelt-verkettete Kantenliste zu transformieren. Fassen wir das Verfahren zusammen. Zunächst ändern wir den Algorithmus BEARBEITEEVENTPUNKT wie folgt:

**Algorithmus** BEARBEITEEVENTPUNKT( $v$ )

- ```

...
3. if ( $U(v) \cup C(v) \cup L(v)$  enthält mindestens ein Segment aus  $\mathcal{S}_1$  und mindestens ein Segment aus  $\mathcal{S}_2$ )
   then
4.     Teile die Kanten in  $U(v) \cup C(v) \cup L(v)$  und aktualisiere die Einträge der Knoten- und Kantentabelle
       in  $\mathcal{D}$  wie oben besprochen.
   end if
...
19. Speichere die Halbkante  $h_d$  direkt links von  $v$  zusammen mit  $v$ . ( $h_d$  gehört zu der Kante, die in  $\mathcal{T}$  der linke
    Nachbar des Eventpunktes ist, d.h. zur Kante  $s_l$  in Zeile 10 bzw. 13 in BEARBEITEEVENTPUNKT. Der
    Ursprung von  $h_d$  ist der obere Endpunkt dieser Kante:  $Ursprung(h_d) \prec_Q Ziel(h_d)$ .)

```

Als nächstes geben wir an, wie die Flächeninformationen aktualisiert werden:

**Algorithmus** FLÄCHENUPDATE()

Eingabe: Keine.

Ausgabe: Keine.

- ```

1. Erzeuge einen leeren Flächeneintrag in  $\mathcal{D}$  für die ungebundene Fläche
2. for each (Eventpunkt  $v$  in  $\mathcal{Q}$ ) do
3.   Sei  $U(v)$  die Menge der Halbkanten  $h$ , so daß  $Ursprung(h) = v$  und  $v \prec_Q Ziel(h)$ 
4.   for each (Halbkante  $h \in U(v)$ ,  $innereFläche(h) \notin \mathcal{D}$ ) do
5.     Sei  $H$  der Halbkantenzug, der in  $h$  beginnt.
6.     if ( $H$  ist gegen den Uhrzeigersinn orientiert)
7.       then
8.         Erzeuge eine neue Fläche  $f$  mit  $äußereBegrenzung(f) := h$  und füge  $f$  in  $\mathcal{D}$  ein.
9.         Speichere mit  $f$  die beiden Flächen  $f_i \in \mathcal{S}_i$  und  $f_j \in \mathcal{S}_j$ ,  $i, j = (1, 2)$ ,  $i \neq j$ , die  $f$  enthalten:
           Ist der Ursprung einer Halbkante  $h_k$  aus  $H$  ein Schnittpunkt, dann entspricht  $f$  den beiden
           Flächen  $f_i = innereFläche(h_k)$  und  $f_j = innereFläche(Vorgänger(h_k))$ . Ist kein Ursprung
           einer Halbkante  $h_k$  aus  $H$  ein Schnittpunkt, dann entspricht  $f$  den beiden Flächen  $f_i =
           innereFläche(h)$  und  $f_j =$  die Fläche, in der  $Ursprung(h)$  liegt.
10.        Setze für jede Halbkante  $h_k$  in  $H$   $innereFläche(h_k) := f$  und aktualisiere  $h_k$  in  $\mathcal{D}$ .
       else
11.        Sei  $h'$  eine Halbkante.
12.        Sei  $v_l$  der am weitesten links liegende Knoten in  $H$ .
13.        Sei  $h_d :=$  die Halbkante, die direkt links von  $v_l$  liegt und für die
            $Ursprung(h_d) \prec_Q Ziel(h_d)$  ist.
14.         $H :=$  der Kantenzug, der in  $h_d$  beginnt.
15.         $h' := h$ .
16.        while ( $H$  existiert und  $H$  ist im Uhrzeigersinn orientiert)
            $h' := h_d$ .

```



17.  $v_l :=$  der am weitesten links liegende Knoten in  $H$ .
18.  $h_d :=$  die Halbkante, die direkt links von  $v_l$  liegt und für die  $Ursprung(h_d) \prec_Q Ziel(h_d)$  ist.
19.  $H :=$  der Kantenzug, der in  $h_d$  beginnt.  
**end while**
20. Sei  $f := innereFläche(h')$ .
21. Sei  $H$  der Kantenzug, der in  $h$  beginnt.
22. Füge  $h$  in  $innereBegrenzung(f)$  ein und aktualisiere den Eintrag von  $f$  in  $\mathcal{D}$ .
23. Setze für jede Halbkante  $h$  in  $H$   $innereFläche(h) := f$  und aktualisiere  $h$  in  $\mathcal{D}$ .  
**end if**  
**end for**  
**end for**

Fügen wir jetzt alles zusammen, kommen wir zu folgendem Algorithmus:

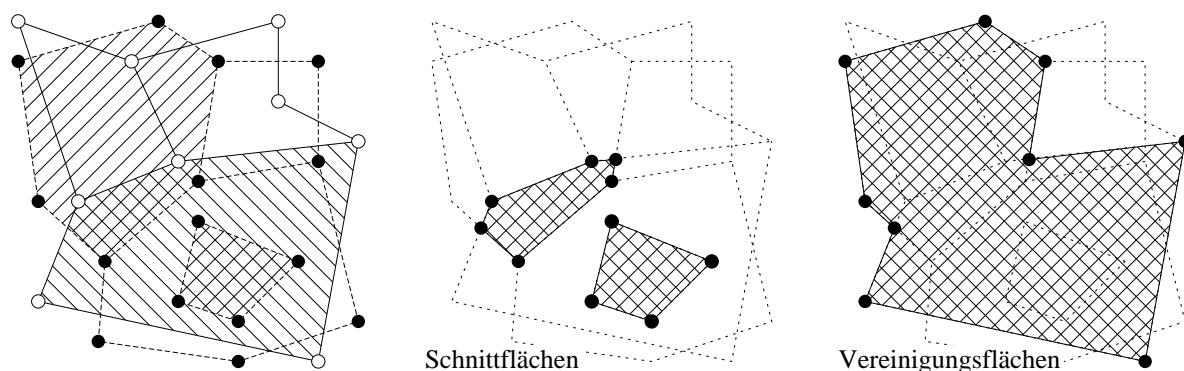
**Algorithmus** KARTENVERSCHNEIDUNG( $\mathcal{S}_1, \mathcal{S}_2$ )

*Eingabe:* Die doppelt-verketteten Kantenlisten zweier planarer Subdivisionen  $\mathcal{S}_1$  und  $\mathcal{S}_2$ .

*Ausgabe:* Das Ergebnis der Verschneidung von  $\mathcal{S}_1$  und  $\mathcal{S}_2$  in der doppelt-verketteten Kantenliste  $\mathcal{D}$ .

1. Kopiere die doppelt-verketteten Kantenlisten von  $\mathcal{S}_1$  und  $\mathcal{S}_2$  in eine neue doppelt-verkettete Kantenliste  $\mathcal{D}$
2. SCHNITTPUNKTSUCHE(alle Kanten aus  $\mathcal{S}_1$  und  $\mathcal{S}_2$ )
3. FLÄCHENUPDATE().

**Satz 3.2.** Sei  $\mathcal{S}_1$  eine planare Subdivision der Komplexität  $n_1$ , sei  $\mathcal{S}_2$  eine planare Subdivision der Komplexität  $n_2$  und sei  $n := n_1 + n_2$ . Die zur Berechnung der Verschneidung von  $\mathcal{S}_1 \cup \mathcal{S}_2$  benötigte Zeit liegt in  $O(n \log n + k \log n)$ , wobei  $k$  die Komplexität der Verschneidung ist.



**Abbildung 3.13** Themabasierte Verschneidungen mit Hilfe Boolescher Junktoren. Die Verschneidung zweier thematischer Karten (links) führt je nach gewähltem logischen Junktoren (Mitte: ‘und’, rechts: ‘oder’) zu geometrisch unterschiedlichen Ergebnissen.

### 3.2.4 Berechnung von Sekundärdaten

Abgesehen von der algorithmischen Seite des Kartenverschneidungsproblems gilt das primäre Interesse in der Regel dem Ergebnis der Verschneidung, insbesondere der Darstellung thematischer Daten auf Grundlage der neu erzeugten Geometriedaten. Eine besondere Bedeutung wird dabei Booleschen Operationen zugemessen, d.h. der Bestimmung von Schnitt-, Vereinigungs- oder Differenzflächen (z.B. Laurini, Thompson 1995, S. 273ff.; de Berg et al. 1997, S. 39f.). Dies zeigt auch unser einführendes Beispiel (vgl. S. 52ff.): Wir hatten angenommen, uns ständen zwei digitale Karten mit thematischen Informationen zur Verfügung – eine Karte der Flächen gleichen Bodentyps und eine der Flurstücke, sowie zwei jeweils eindeutig referenzierte Tabellen mit den thematischen Attributen ‘Bodentyp’ und ‘Bodennutzung’ – und haben die Frage gestellt nach den Schwarzerdeflächen, auf denen Zuckerrüben angebaut werden. Wir hatten weiter angenommen, die beiden Karten seien nicht grenzscharf, so daß die Antwort auf die Frage nach den Flächen  $f$  mit dem thematischen Attribut ‘Schwarzerde  $\wedge$  Zuckerrüben’ nicht unmittelbar aus den beiden gegebenen Karten ableitbar, sondern nur über den ‘Umweg’ der geometrischen Verschneidung zu finden war: Gesucht waren die Schnittflächen aller Flächen der Bodentypenkarte mit dem Attribut ‘Schwarzerde’ und aller Flächen der Bodennutzungskarte mit dem Attribut ‘Zuckerrüben’. Ist nun die Verschneidung  $Sub(Bodentypkarte, Bodennutzungskarte)$  bekannt, können die gesuchten Flächen daraus extrahiert werden, denn dann kennen wir ja für jede Fläche der Verschneidung auch die beiden ‘Originalflächen’ (je eine Fläche gleichen Bodentyps und eine Fläche gleicher Bodennutzung), aus denen die Schnittfläche hervorgegangen ist.

Im Fall unserer beiden Beispielkarten heißt das, daß wir nur jene Flächen tatsächlich in die doppelt-verkettete Kantenliste  $\mathcal{D}$  zurückschreiben müssen, deren ‘Originalflächen’ thematisch dem Attribut ‘Schwarzerde  $\wedge$  Zuckerrüben’ entsprechen (Schritte 7-9 in FLÄCHENUPDATE). Hätten wir nach den Flächen gefragt, auf denen entweder Schwarzerde vorkommt oder Zuckerrüben wachsen, wäre die Lage etwas anders, zumindest dann, wenn wir die Flächen mit ‘Schwarzerde  $\vee$  Zuckerrüben’ *geometrisch korrekt* bestimmen wollen. Abb. 3.13 verdeutlicht das Problem (vgl. auch Abb. 3.2): Bei der Bestimmung von Vereinigungsflächen müssen die gemeinsamen Umrisse derjenigen Schnittflächen bestimmt werden, für die mindestens eines der beiden Attribute zutrifft. Die doppelt-verkettete Kantenliste  $\mathcal{D}$  enthält dementsprechend auch nicht alle Knoten, Kanten und Schnittpunkte aus  $S_1 \cup S_2$ . Unabhängig von den gewünschten Eigenschaften der ‘Ergebnisflächen’ müssen wir aber in jedem Fall zunächst alle Schnittflächen bestimmen; anders können wir im Einzelfall nicht entscheiden, ob eine Schnittflä-

che den gewünschten Attributen entspricht oder ob nicht.<sup>16</sup> Mit anderen Worten: Das Verfahren muß in jedem Fall vollständig durchgeführt werden, und die Laufzeit hängt in keiner Weise von unserer inhaltlichen Fragestellung ab.

Es ist nicht unbedingt zwingend notwendig, Vereinigungs- und Differenzflächen geometrisch korrekt, d.h. ohne überflüssige Knoten und Kanten zu berechnen. Lassen wir den Speicherbedarf für  $\mathcal{D}$  (der natürlich linear von der Komplexität der darin gespeicherten planaren Subdivision abhängig ist) unberücksichtigt, können wir – ohne zusätzliche Laufzeit in Kauf nehmen zu müssen – ebensogut alle Flächen der Verschneidung in  $\mathcal{D}$  speichern und die Zeiger auf die ‘Originalflächen’ als deren thematische Attribute betrachten. Die Durchführung der Booleschen Operation ‘Schwarzerde  $\wedge$  Zuckerrüben’ entspricht dann der Aufgabe, für die Flächen der Verschneidung zunächst das Sekundärdatum (den Vektor)  $x = \text{Schwarzerde} \wedge \text{Zuckerrüben}$  und anschließend die Teilmenge derjenigen Flächen zu bestimmen, für die der Wert von  $x$  ‘wahr’ ist. Diese Operationen können für beliebige Junktoren, also neben ‘ $\wedge$ ’ auch für ‘ $\vee$ ’ oder ‘ $\setminus$ ’, mit Hilfe des Datenbanksystems durchgeführt werden, das die thematischen Daten vorhält. Dann ist für jede Fläche in  $f \in \text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$ , für die die beiden Flächen  $f_1 \in \mathcal{S}_1$  und  $f_2 \in \mathcal{S}_2$ , die  $f$  enthalten, bekannt sind, die Berechnung von Sekundärdaten für  $f$  aufgrund der thematischen Daten von  $f_1$  und  $f_2$  trivial. In diesem Sinne ist die Bestimmung von Schnitt-, Vereinigungs- oder Differenzflächen dann kein geometrisches Problem, sondern ist dem geometrischen Problem der Bestimmung von  $\mathcal{D}$  nachgeordnet.

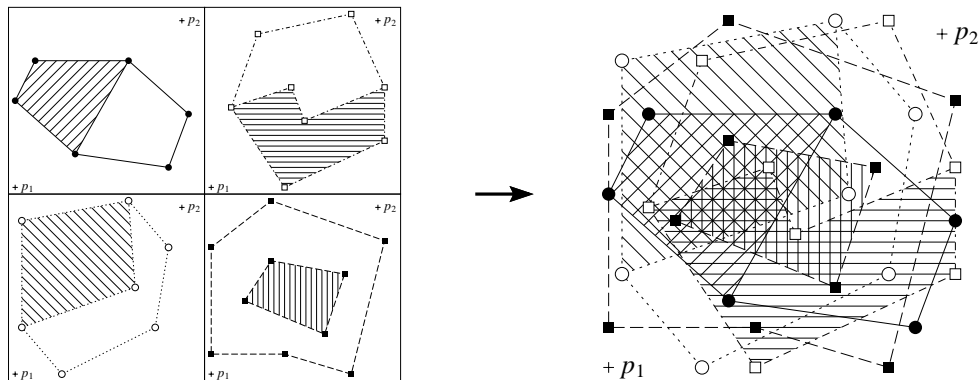
### 3.3 Die Verschneidung von $m$ Flächenkarten

#### 3.3.1 Problemstellung

Nachdem wir das ‘klassische’ Kartenverschneidungsproblem (vgl. S. 54) gelöst haben, wollen wir nun versuchen, das Verfahren zu verallgemeinern.<sup>17</sup> Nehmen wir an, für unseren fiktiven Untersuchungsraum stände uns neben den beiden Karten des Bodentyps und der Bodennutzung eine dritte Karte zur Verfügung, die uns Auskunft über die lokalen Temperaturverhältnisse gibt. Auch diese dritte Karte soll als planare Subdivision der Untersuchungsfläche vorliegen, z.B. in Form von durch Isothermen begrenzten Flächen. Auf Grundlage der Karteninformationen könnten wir nun beispielsweise fragen, wo (1.) Schwarzerdeflächen vorkommen, auf denen (2.) Zuckerrüben wachsen und auf denen (3.) die durchschnittliche Tageshöchsttemperatur im März über  $15^\circ$  liegt. Die Lösung des Problems ist einfach: Wir bestimmen die Schnittflächen der ersten beiden Karten mit  $(\text{Bodentyp} = \text{Schwarzerde}) \wedge (\text{Bodennutzung} = \text{Zuckerrüben})$  und verschneiden das Ergebnis mit der dritten, d.h. wir bestimmen die Schnittflächen mit  $((\text{Bodentyp} = \text{Schwarzerde}) \wedge (\text{Bodennutzung} = \text{Zuckerrüben})) \wedge (\text{Tageshöchsttemperatur im März} > 15^\circ)$ . Es fällt nicht schwer, sich vorzustellen, daß eine vierte als planare Subdivision vorliegende Karte (z.B. eine Karte der Niederschlagsverteilung) mit dem Ergebnis der Verschneidung und das Ergebnis dieser Verschneidung wiederum mit einer fünften Karte (z.B. einer Karte der Höhen-

<sup>16</sup>Für die Bestimmung von Vereinigungsflächen durchqueren wir nach Abschluß von FLÄCHENUPDATE einfach den der planaren Subdivision  $\text{Sub}(\mathcal{S}_1, \mathcal{S}_2)$  zugrundeliegenden Graph, um die entsprechenden Flächen ‘einzusammeln’. Wir starten dabei jeweils auf einer Halbkante, deren Ursprung der am weitesten links liegende Punkt aller benachbarter Schnittflächen ist, und folgen dem Kantenzug, indem wir an jedem Knoten unter Berücksichtigung der Attributwerte möglichst weit rechts abbiegen. Die gefundenen Kantenzüge entsprechen dann den Begrenzungen der gesuchten Vereinigungsflächen, und nur diese werden inclusive ihrer Knoten und Kanten in der doppelt-verketteten Kantenliste  $\mathcal{D}$  gespeichert (zuvor müssen wir natürlich eine temporäre doppelt-verkettete Kantenliste  $\mathcal{D}'$  benutzen). Ähnlich ist auch das Problem der Bestimmung von Differenzflächen gelagert; man beachte aber, daß der Differenzjunktoren nicht reflexiv ist! Das Durchqueren des Graphen erfordert zusätzlich  $O(n + k)$  Zeit.

<sup>17</sup>Die Ausführungen in Kap. 3.2 bezogen sich überwiegend auf die Darstellung von de Berg et al. 1997, 19ff. Die Verallgemeinerung des Kartenverschneidungsproblems ist dagegen nicht Gegenstand weder der Geographie noch der Geoinformatik, und die Anzahl der ohnehin nicht sehr zahlreich zur Verfügung stehenden Literaturquellen geht nun gegen Null.



**Abbildung 3.14** Eine visuelle Verschneidung von vier Karten. Auch vier durch mindestens zwei Punkte  $p_1$  und  $p_2$  in der Euklidischen Ebene referenzierte thematische Karten (links) können geometrisch korrekt übereinandergelegt und so ‘visuell verschnitten’ werden (rechts). Die Schnittflächen sind wiederum sichtbar, jedoch in ihren geometrischen Eigenschaften unbekannt. Die Überlagerung entspricht dem Inhalt der doppelt-verketteten Kantenliste  $\mathcal{D}$  vor der Verschneidung.

lagen ü.N.N.) verschnitten werden kann. Alle Flächen des Untersuchungsraums mit der Eigenschaft  $((\text{Bodentyp} = \text{Schwarzerde}) \wedge (\text{Bodennutzung} = \text{Zuckerrüben})) \wedge (\text{Tageshöchsttemperatur im März} > 15^\circ) \vee ((\text{Mittlerer Niederschlag im März} < 100\text{mm}) \wedge (\text{Höhenlage} > 400\text{mm}))$  können z.B. mit Hilfe von vier einzelnen Verschneidungen bestimmt werden.

Für den Fall, daß mehr als zwei Flächenkarten verschnitten werden sollen, müssen wir also einer ‘paarweisen Verschneidestrategie’ folgen, indem wir mit der Verschneidung von zwei Karten beginnen und das Ergebnis mit der dritten Flächenkarte, dieses Ergebnis mit der vierten usw. verschneiden, solange, bis alle Flächenkarten verarbeitet worden sind. Entsprechend erfordert die Verschneidung von  $m$  planaren Subdivisionen  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m$ <sup>18</sup> genau  $m - 1$  einzelne Verschneidungen. Im folgenden wird gezeigt, wie das Problem wesentlich effektiver gelöst werden kann. Zunächst formulieren wir die Aufgabenstellung präzise als

**Allgemeines Problem der Flächenverschneidung:** Die Verschneidung der  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  ist die planare Subdivision  $Sub(\mathcal{S}_{1\dots m})$ , für die eine Fläche  $f$  genau dann existiert, wenn es die Flächen  $f_1 \in \mathcal{S}_1, f_2 \in \mathcal{S}_2, \dots, f_m \in \mathcal{S}_m$  gibt, so daß  $f$  eine maximal verbundene Teilmenge aus  $f_1 \cap f_2 \cap \dots \cap f_m$  ist. Bestimme  $Sub(\mathcal{S}_{1\dots m})$ .

Für  $Sub(\mathcal{S}_{1\dots m})$  verlangen wir natürlich, daß für jede Fläche  $f \in Sub(\mathcal{S}_{1\dots m})$  die Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  bekannt sind, die  $f$  enthalten. Auf diese Weise können wir für jede Fläche  $f \in Sub(\mathcal{S}_{1\dots m})$  – ganz wie bei dem in Kap. 3.2 erörterten Fall für  $m = 2$  – auf die thematischen Daten der ‘Originalflächen’ zurückgreifen. Abb. 3.14 zeigt ein Beispiel eines allgemeinen Flächenverschneidungsproblems für  $m = 4$ . Blenden wir in Gedanken zunächst die thematischen Flächeninformationen aus: Übrig bleiben Kanten, von denen sich einige überschneiden, und Knoten, von denen einige auf Kanten liegen. Genau wie beim zweifachen Flächenverschneidungsproblem wird die Euklidische Ebene in neue Teilflächen partitioniert, die zwar unmittelbar sichtbar sind, deren Grenzen aber in ihren geometrischen und topologischen Eigenschaften unbekannt sind. Analog zum Problem der Kartenverschneidung muß zur Lösung des allgemeinen Problems der Flächenverschneidung daher zunächst diejenige planare Subdivision bestimmt werden, die durch die Knoten und Kanten in  $\bigcup_{i=1}^m \mathcal{S}_i$  induziert wird. Damit ist bereits klar,

<sup>18</sup>Im folgenden kurz  $\mathcal{S}_{1\dots m}$

welche Verfahrensschritte wir aus Kap. 3.2 übernehmen können: Zunächst kopieren wir die doppelt-verketteten Kantenlisten aller planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  in eine neue doppelt-verkettete Kantenliste  $\mathcal{D}$ , die dann ungültig ist, da sich (aller Voraussicht nach) Kanten überschneiden (Abb. 3.14). Die Aufgabe des gesuchten Algorithmus besteht wiederum darin, die ungültige doppelt-verkettete Kantenliste  $\mathcal{D}$  in eine gültige doppelt-verkettete Kantenliste zu transformieren. Diese Transformation erfordert die Bestimmung aller Schnittpunkte der Kanten aus  $\bigcup_{i=1}^m \mathcal{S}_i$  und die entsprechenden Aktualisierungen der Knoten und Halbkanten in  $\mathcal{D}$ . Die Bestimmung des der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$  zugrundeliegenden Graphen erfolgt also genauso wie in Kap. 3.2.3.1 besprochen. Entsprechend erfolgt die Zuordnung von äußeren und inneren Flächengrenzen nach den Regeln aus Kap. 3.2.3.2. Unabhängig davon, ob ursprünglich zwei, vier oder  $m$  planare Subdivisionen ‘übereinandergelegt’ wurden, ist das Ergebnis der Verschneidung wieder *eine* planare Subdivision, und für die doppelt-verkettete Kantenliste  $\mathcal{D}$  der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$  gelten dieselben Regeln wie für alle bisher besprochenen doppelt-verketteten Kantenlisten.

Anders ist die Situation bei der Aktualisierung der thematischen Informationen. In Kap. 3.2.3.3 haben wir gesehen, daß das Problem darin besteht, den ‘Ergebnisflächen’ der Verschneidung als Label die ‘Originalflächen’ zuzuordnen, die sie enthalten. Bei der Verschneidung von zwei planaren Subdivisionen  $\mathcal{S}_i$  und  $\mathcal{S}_j$  und unter Berücksichtigung der ungebundenen Flächen besteht jedes Label strenggenommen aus genau zwei Flächen, nämlich je eine aus  $\mathcal{S}_i$  und eine aus  $\mathcal{S}_j$ .<sup>19</sup> Da wir an den ungebundenen Flächen in der Regel nicht interessiert sind (weil dort ja ‘nichts’ ist), bestehen die Label in der Praxis aus höchstens zwei ‘Originalflächen’.

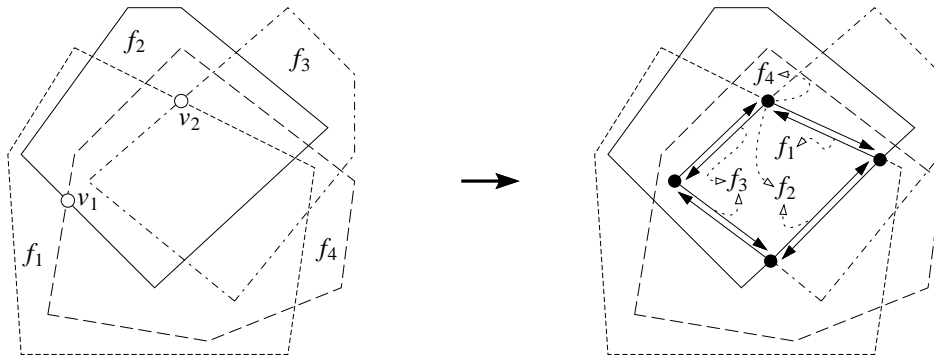
Bei den Flächen, deren äußerer Kantenzug einen Schnittpunkt enthält, konnten die beiden Originalflächen direkt an den *innereFläche()*-Zeigern abgelesen werden. In allen anderen Fällen mußte für mindestens einen Knoten  $v \in \mathcal{S}_i$  auf dem äußeren Kantenzug bekannt sein, innerhalb welcher Fläche  $f \in \mathcal{S}_j$  ( $i \neq j$ ) er liegt (Kap. 3.2.3.3). Beim Flächenverschneidungsproblem für  $m = 2$  gilt also folgende Fallunterscheidung: Jeder Knoten ist entweder ein *Schnittpunkt*, der auf zwei ursprünglichen Flächengrenzen (auf je einer aus  $\mathcal{S}_i$  und  $\mathcal{S}_j$ ) und niemals innerhalb einer ‘Originalfläche’ liegt, oder er ist *kein Schnittpunkt*, so daß er auf genau einer ursprünglichen Flächengrenze aus  $\mathcal{S}_i$  liegt und innerhalb einer Fläche aus  $\mathcal{S}_j$  ( $i \neq j$ ) liegen kann. Abb. 3.15 (links) zeigt, daß diese Unterscheidung für  $m > 2$  nicht mehr zutrifft: Eine Flächengrenze aus  $\mathcal{S}_i$  kann eine Flächengrenze aus  $\mathcal{S}_j$  so schneiden, daß der Schnittpunkt innerhalb einer Fläche liegt, die zu einer dritten planaren Subdivision  $\mathcal{S}_k$  gehört (wie z.B.  $v_1$ ); für  $m > 3$  kann der Schnittpunkt auch innerhalb mehrerer Flächen liegen (wie z.B.  $v_2$ ).

### 3.3.2 Bestimmung der Flächenlabel

Betrachten wir ganz allgemein einen Knoten  $v$  der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$ , der durch die Überschneidung einer Flächengrenze aus  $\mathcal{S}_i$  und einer Flächengrenze aus  $\mathcal{S}_j$  entsteht ( $i < j$ ). Der Knoten  $v$  liegt natürlich niemals innerhalb einer Fläche aus  $\mathcal{S}_i$  oder einer Fläche aus  $\mathcal{S}_j$ . Es ist aber möglich, daß  $v$  innerhalb je einer Fläche aus  $\mathcal{S}_1, \dots, \mathcal{S}_{i-1}, \mathcal{S}_{i+1}, \dots, \mathcal{S}_{j-1}, \mathcal{S}_{j+1}, \dots, \mathcal{S}_m$  liegt. Jeder Knoten der Verschneidung  $Sub(\mathcal{S}_{1\dots m})$ , in dem sich zwei Flächengrenzen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  schneiden, liegt also innerhalb von höchstens  $m - 2$  Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ . Entsprechend liegt jeder Knoten aus  $Sub(\mathcal{S}_{1\dots m})$ , in dem sich  $r$  ( $r \leq m$ ) Flächengrenzen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  schneiden, innerhalb von höchstens  $m - r$  Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ .<sup>20</sup> Anders als beim Flächenverschneidungsproblem für  $m = 2$  kann die Frage, welche Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$  eine Fläche  $f$  aus  $Sub(\mathcal{S}_{1\dots m})$  enthalten, daher nicht mehr lokal an einem Knoten der äußeren Begrenzung von  $f$  entschieden werden. Wir müssen vielmehr die äußere Begrenzung von

<sup>19</sup>Zur Verdeutlichung: Das Label der ungebundenen Fläche der Verschneidung  $Sub(\mathcal{S}_1, \mathcal{S}_2)$  besteht in diesem Sinne aus der ungebundenen Fläche von  $\mathcal{S}_1$  und der ungebundenen Fläche von  $\mathcal{S}_2$ .

<sup>20</sup>Berücksichtigt man die ungebundenen Flächen, liegt *jeder* Knoten aus  $Sub(\mathcal{S}_{1\dots m})$ , in dem sich  $r$  ( $r \leq m$ ) Flächengrenzen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  schneiden, natürlich in *genau*  $m - r$  Flächen.



**Abbildung 3.15** Die Bestimmung der Flächenlabel für  $m > 2$ . Wenn sich Flächen aus mehr als zwei planaren Subdivisionen überlappen, können Schnittpunkte innerhalb anderer Flächen liegen (links). Um das Label einer Fläche der Verschneidung zu bestimmen, muß die Fläche vollständig umlaufen werden (rechts).

$f$  vollständig umlaufen und dabei die Flächenlabel ‘einsammeln’ (vgl. Abb. 3.15 rechts): Sei  $H$  der Halbkantenzug, der die Fläche  $f$  nach außen begrenzt, d.h. der in *äußereBegrenzung*( $f$ ) beginnt. Das Label der Fläche  $f$  ist die Vereinigungsmenge aller Flächen, die durch die *innereFläche*( $h$ )-Zeiger jeder Halbkante  $h$  aus  $H$  referenziert sind, und aller Flächen, in denen die Knoten *Ursprung*( $h$ ) jeder Halbkante  $h$  aus  $H$  liegen. Das einmalige Umlaufen jedes eine Fläche in  $Sub(\mathcal{S}_{1\dots m})$  begrenzenden Kantenzuges erfordert eine Zeit, die linear zur Länge des Kantenzuges ist.

### 3.3.3 Bestimmung der umgebenden Flächen

Für jede Halbkante in  $\mathcal{D}$  ist die innere Fläche bekannt, so daß alle Flächen, die durch die *innereFläche*( $\cdot$ )-Zeiger referenziert sind, beim Umlaufen einer Fläche  $f$  gefunden werden. Wie aber wissen wir für jeden Knoten  $v$ , welche Flächen aus  $\mathcal{S}_{1\dots m}$  ihn enthalten? Bei der Verschneidung von zwei planaren Subdivisionen war die Idee die, daß es für jeden Knoten  $v \in \mathcal{S}_i$ , der innerhalb einer Fläche aus  $\mathcal{S}_j$  ( $i \neq j$ ) liegt, eine Kante aus  $\mathcal{S}_j$  geben muß, die links von  $v$  liegt. Da die links benachbarte Kante  $e_l$  für jeden Knoten  $v$  bekannt ist, ließ sich die gesuchte Flächengrenze schnell ermitteln: Gehörte  $e_l$  zu  $\mathcal{S}_i$ , dann lag  $v$  in der Fläche, auf die der *innereFläche*( $\cdot$ )-Zeiger derjenigen Halbkante  $h_d$  von  $e_l$  zeigte, deren Ursprung gleich dem oberen Endpunkt von  $e_l$  war. Gehörte  $e_l$  dagegen zu  $\mathcal{S}_j$ , also zur selben planaren Subdivision wie  $v$ , dann lag  $v$  genau dann innerhalb derselben Fläche wie *Ursprung*( $h_d$ ), wenn  $v$  nicht auf der Begrenzung dieser Fläche lag (vgl. Kap. 3.2.3.3). Wir nutzten dabei aus, daß für einen beliebigen Knoten, der nicht ein Schnittpunkt ist, und der Kante, die zu diesem Knoten direkt links benachbart ist, die Zugehörigkeit zu einer der beiden planaren Subdivisionen  $\mathcal{S}_i$  oder  $\mathcal{S}_j$  bekannt ist. Bei der Verschneidung von  $m > 2$  planaren Subdivisionen kann nun jeder Knoten  $v$  in maximal  $m - 1$  Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  liegen. Wir wollen hier auf die Kenntnis der Zugehörigkeit von Knoten und Kanten zu einer der planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  verzichten<sup>21</sup> und das Problem der Flächenlabel rein topologisch lösen. Dazu formulieren wir das

<sup>21</sup>Der Grund dafür ist einfach: Wenn wir aus Daten geometrische Figuren konstruieren, legen wir diese auf *einen* Layer; wir erzeugen also im Prinzip *eine* Subdivision, deren zugrundeliegender Graph in der Regel nicht planar ist. Um bei unserer Diskussion nicht unnötig zwischen geometrischen Figuren, Karten usf. differenzieren zu müssen, fassen wir jede einzelne geometrische Figur als eigene planare Subdivision auf. In der Praxis führt dies jedoch zu nichts weiter als einer doppelten Indizierung, je eine für die Figur selbst und eine für die planare Subdivision. Der Einfachheit halber verzichten wir daher im folgenden grundsätzlich auf die Kenntnis der Zugehörigkeit von Knoten, Kanten und Flächen zu einer der ‘ursprünglichen’ planaren Subdivisionen  $\mathcal{S}_{1\dots m}$ .

**Problem der umgebenden Flächen:** Gegeben sei eine planare Subdivision  $Sub(\mathcal{S}_{1\dots m})$  als vorläufiges Ergebnis der Verschneidung der  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  und die dazugehörige doppelt-verkettete Kantenliste  $\mathcal{D}$ . In  $\mathcal{S}_{1\dots m}$  gebe es keine zwei Kanten, die sich überlappen. In  $\mathcal{D}$  seien die Knoten- und Flächentabelle vollständig, die Kantentabelle bis auf die *innereFläche*()-Einträge vollständig aktualisiert. Bestimme für einen beliebigen Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  die Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die ihn enthalten.

Die umschließende Fläche eines Knotens für  $m = 2$  wurde nach der Beschreibung in Kap. 3.2.3.3 während der Aktualisierung der Flächeneinträge in der doppelt-verketteten Kantenliste  $\mathcal{D}$  bestimmt. Diese Strategie soll auch im folgenden beibehalten werden; entscheidend ist, daß das Problem der umgebenden Flächen während eines Durchlaufs der Eventfolge  $\mathcal{Q}$  bearbeitet wird. Überlegen wir uns zunächst, was wir mit Sicherheit wissen (vgl. dazu Abb. 3.16):

- Ein Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  liegt niemals innerhalb einer Fläche  $f \in \{\bigcup_{i=1}^m \mathcal{S}_i\}$ , wenn  $v$  auf der Grenze von  $f$  liegt.
- Liegt ein Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  innerhalb einer Fläche  $f \in \{\bigcup_{i=1}^m \mathcal{S}_i\}$ , dann gibt es sowohl links als auch rechts von  $v$  eine Kante, die  $f$  begrenzt.
- Liegt ein Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  innerhalb einer Fläche  $f \in \{\bigcup_{i=1}^m \mathcal{S}_i\}$ , dann verläuft der Halbkantenzug, der  $f$  begrenzt, von  $v$  aus gesehen gegen den Uhrzeigersinn.
- Gibt es eine Kante  $e_l \in Sub(\mathcal{S}_{1\dots m})$ , die zu einem Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  in der Statusfolge  $\mathcal{T}$  links benachbart ist, wenn  $v$  der aktuelle Eventpunkt in der Eventfolge  $\mathcal{Q}$  ist, dann gibt es einen Knoten  $v_u$ , der der obere Endpunkt von  $e_l$  ist und in der Eventfolge  $\mathcal{Q}$  vor  $v$  steht.

All diese Überlegungen sind uns bereits geläufig. Die letzte Bemerkung verdient nun unsere besondere Aufmerksamkeit, denn sie bildet die Grundlage für unser Verfahren der Flächenlabelaktualisierung. Die Idee ist wie folgt (vgl. Abb. 3.16): Wenn zu einem Knoten  $v$  der Knoten  $v_u$  existiert, der der obere Endpunkt der links von  $v$  benachbarten Kante ist, dann kann zwischen  $v$  und  $v_u$  keine weitere Kante existieren, die zwei Flächen so voneinander trennt, daß  $v$  in der einen und  $v_u$  in der anderen Fläche liegt. Daher können wir die ‘Originalflächen’ aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die einen Knoten  $v \in Sub(\mathcal{S}_{1\dots m})$  enthalten, im Prinzip aufgrund der ‘Originalflächen’ bestimmen, die den Knoten  $v_u$  enthalten.

**Hilfssatz 3.3.** *Sei  $v$  ein Knoten der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$ , sei  $e_l \in Sub(\mathcal{S}_{1\dots m})$  die Kante, die zu  $v$  in der Statusfolge  $\mathcal{T}$  links benachbart ist, wenn  $v$  der aktuelle Eventpunkt in der Eventfolge  $\mathcal{Q}$  ist, und sei  $v_u$  der obere Endpunkt von  $e_l$ . Dann gibt es keine zwei Flächen  $f_1 \in \{\bigcup_{i=1}^m \mathcal{S}_i\}$  und  $f_2 \in \{\bigcup_{i=1}^m \mathcal{S}_i\}$ , so daß weder  $v_u$  noch  $v$  auf einer der Begrenzungen von  $f_1$  oder  $f_2$  liegen und  $v_u \in f_1$  und  $v \in f_2$ .*

*Beweis.* Da die Kante  $e_l$  und der Knoten  $v$  in  $\mathcal{T}$  benachbart sind, gibt es keinen Punkt  $p \in Sub(\mathcal{S}_{1\dots m})$ , der auf einer Kante aus  $Sub(\mathcal{S}_{1\dots m})$  liegt und nicht der untere Endpunkt dieser Kante ist, so daß  $p.y = v.y$  und  $e_l \prec_{\mathcal{T}} p \prec_{\mathcal{T}} v$ . Daher begrenzen  $e_l$  und  $v$  dieselbe Fläche. Da  $v_u$  an  $e_l$  gebunden ist, begrenzen auch  $v_u$  und  $v$  dieselbe Fläche. Sei  $f$  die Fläche aus  $Sub(\mathcal{S}_{1\dots m})$ , auf deren Grenze  $v_u$  und  $v$  liegen. Dann gibt es einen verbundenen Kantenzug  $E$ , dessen Knoten und Kanten vollständig in  $f$  liegen und der  $v$  mit  $v_u$  verbindet.

Sei nun  $f_1$  eine Fläche in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v_u$  enthält,  $v$  nicht enthält und deren Grenze nicht durch  $v$  verläuft, und sei  $f_2$  eine Fläche in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v$  enthält,  $v_u$  nicht enthält und deren Grenze nicht durch  $v_u$  verläuft. Die Begrenzungen sowohl von  $f_1$  als auch von  $f_2$  müssen dann so zwischen  $v$  und  $v_u$  verlaufen, daß sie den Kantenzug  $E$  schneiden. Das bedeutet aber, daß die Begrenzungen von  $f_1$

und  $f_2$  jeweils zum Teil innerhalb von  $f$  liegen. Da die planare Subdivision  $Sub(\mathcal{S}_{1\dots m})$  das Ergebnis der Verschneidung von  $\bigcup_{i=1}^m \mathcal{S}_i$  darstellt, müssen die Begrenzungen von  $f_1$  und  $f_2$  auch Knoten und Kanten in  $Sub(\mathcal{S}_{1\dots m})$  sein. Das heißt wiederum, daß Knoten und/oder Kanten aus  $Sub(\mathcal{S}_{1\dots m})$  innerhalb einer Fläche aus  $Sub(\mathcal{S}_{1\dots m})$  liegen müssen. Dies ist unmöglich, da  $Sub(\mathcal{S}_{1\dots m})$  planar ist. Daher existieren keine zwei Flächen  $f_1 \in Sub(\mathcal{S}_{1\dots m})$  und  $f_2 \in Sub(\mathcal{S}_{1\dots m})$ , so daß weder  $v_u$  noch  $v$  auf einer der Begrenzungen von  $f_1$  oder  $f_2$  liegen und  $v_u \in f_1$  und  $v \in f_2$ .  $\square$

Hilfssatz 3.3 ist uns im Prinzip bereits aus Kap. 3.2.3.3 vertraut, wo er ohne Beweis vorausgesetzt wurde. Er führt uns nun zu folgender Vorgehensweise: Wenn keine links von einem beliebigen Eventpunkt  $v \in Sub(\mathcal{S}_{1\dots m})$  benachbarte Kante existiert, dann ist  $v$  auch in keiner Fläche aus  $\bigcup_{i=1}^m \mathcal{S}_i$  enthalten. Ansonsten liegt  $v$  innerhalb oder auf der Begrenzung derjenigen Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die einen anderen, in der Eventfolge vor  $v$  gelegenen Eventpunkt  $v_u$  enthalten. Dieser Knoten  $v_u$  ist der obere Endpunkt der links von  $v$  benachbarten Kante. Die Begrenzungen aller anderen Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v$ , nicht aber  $v_u$  enthalten, verlaufen in  $Sub(\mathcal{S}_{1\dots m})$  gegen den Uhrzeigersinn durch  $v_u$ .

Wir machen uns hier erneut die Invariante des Scanline-Prinzips zunutze, nach der links und oberhalb eines Eventpunktes sämtliche Informationen korrekt berechnet worden sind. Wir kennen daher bereits die ‘Originalflächen’, die  $v_u$  enthalten (sie sind mit  $v_u$  gespeichert). Die Flächen, deren Grenzen durch  $v$  verlaufen, können wir ohne Schwierigkeiten lokal am Knoten  $v$  bestimmen. Was bleibt, ist die Bestimmung derjenigen Flächen, deren Grenzen durch  $v_u$  verlaufen und die  $v$  enthalten. Bevor wir damit fortfahren, soll jedoch zunächst die Richtigkeit unserer Vorgehensweise nachgewiesen werden:

**Satz 3.4.** *Sei  $v$  ein Knoten der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$ , und sei  $v_u$  der obere Endpunkt der zu  $v$  in der Statusfolge  $\mathcal{T}$  links benachbarten Kante. Sei  $E(v_u)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v_u$  enthalten, sei  $L(v_u)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , deren äußere Grenzen durch  $v_u$  verlaufen und gegen den Uhrzeigersinn orientiert sind, und sei  $B(v)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , deren Grenzen durch  $v$  verlaufen. Dann ist die Menge aller Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v$  enthalten,  $E(v) = \{E(v_u) \cup L(v_u)\} \setminus B(v)$ .*

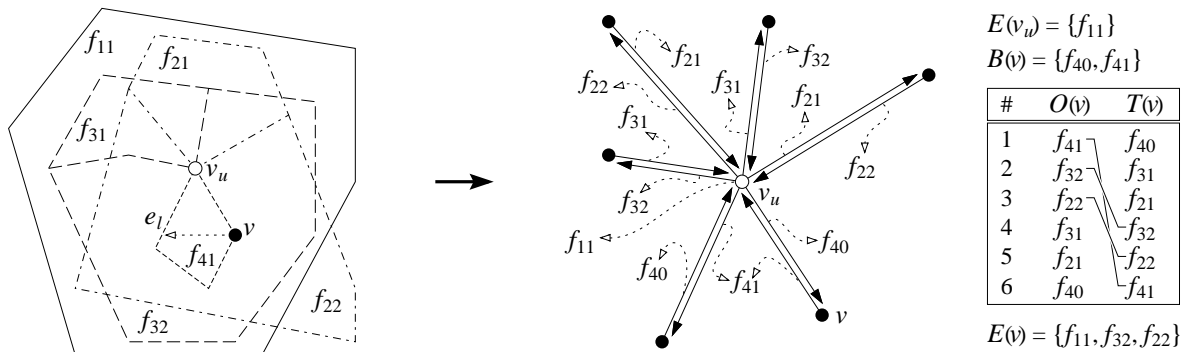
*Beweis.* Für jede Fläche  $f \in E(v_u)$  gilt wegen Hilfssatz 3.3, daß  $f$  entweder  $v$  enthält oder daß  $v$  auf der Begrenzung von  $f$  liegt. Wegen der Invariante gibt es neben den Flächen in  $E(v_u)$  keine weiteren Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v_u$  enthalten. Daher sind mit  $E(v_u)$  auch ausnahmslos alle Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  bekannt, die sowohl  $v_u$  als auch  $v$  enthalten. Zusätzlich zu den Flächen in  $E(v_u)$  kann  $v$  also nur in solchen Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  enthalten sein, die  $v_u$  nicht enthalten.

Für jede Fläche  $f \in \bigcup_{i=1}^m \mathcal{S}_i$ , die  $v$  enthält, gilt wegen Hilfssatz 3.3 wiederum, daß  $f$  entweder  $v_u$  enthält oder daß  $v_u$  auf der Begrenzung von  $f$  liegt. Da alle Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die sowohl  $v_u$  als auch  $v$  enthalten, mit  $E(v_u)$  bekannt sind, verlaufen die Grenzen aller Flächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$ , die nicht  $v_u$ , wohl aber  $v$  enthalten, durch  $v_u$ .

Da die Begrenzung jeder Fläche, die  $v$  enthält, per definitionem gegen den Uhrzeigersinn orientiert sein muß, enthält  $L(v_u)$  genau die Flächen, die nicht  $v_u$ , wohl aber  $v$  enthalten. Neben den Flächen in  $E(v_u) \cup L(v_u)$  kann es also keine weiteren Flächen geben, die  $v$  enthalten. Zusätzlich zu den gesuchten Flächen können aber sowohl  $E(v_u)$  als auch  $L(v_u)$  solche Flächen enthalten, deren Grenzen durch  $v$  verlaufen. Die Menge  $B(v)$  dieser Flächen kann anhand aller an  $v$  gebundenen Kanten bestimmt werden. Die Flächen in  $B(v)$  enthalten  $v$  nicht. Daher ist die Menge aller Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v$  enthalten, gleich  $\{E(v_u) \cup L(v_u)\} \setminus B(v)$ .  $\square$

Ohne Probleme finden wir immerhin einen Teil der Flächen, die  $v$  enthalten: In Abb. 3.16 links sind – der Terminologie von Satz 3.4 entsprechend –  $E(v_u) = \{f_{11}\}$  und  $B(v) = \{f_{40}, f_{41}\}$ .  $E(v_u)$  ist wegen der Invariante bekannt,  $B(v)$  läßt sich leicht ermitteln, wenn  $v$  der aktuelle Eventpunkt ist. Wir wissen daher bereits, daß die Fläche  $f_{11}$  den Knoten  $v$  enthält. Allerdings ist  $v$  noch in zwei





**Abbildung 3.16** Das Problem der umgebenden Flächen. Dargestellt sind vier sich überlappende planare Subdivisionen  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$  und  $\mathcal{S}_4$  (links). Mit  $f_{ij}$  wird die  $j$ -te Fläche der planaren Subdivision  $\mathcal{S}_i$  bezeichnet,  $f_{i0}$  bezeichnet die ungebundene Fläche von  $\mathcal{S}_i$ . Von den vier planaren Subdivisionen bestehen zwei ( $\mathcal{S}_1$  und  $\mathcal{S}_4$ ) aus je einer Fläche, die anderen zwei ( $\mathcal{S}_2$  und  $\mathcal{S}_3$ ) aus je zwei Flächen. Keine dieser Flächen enthält Löcher, so daß im Bereich der Fläche  $f_{41}$  vier verschiedene Ursprungsflächen übereinanderliegen. Dies sind die Flächen  $f_{11}, f_{22}, f_{32}$  und  $f_{41}$ . Um diese Erkenntnis jedoch algorithmisch zu gewinnen, müssen wir zunächst entscheiden, welche Ursprungsflächen den Knoten  $v$  enthalten (Mitte und rechts).

weiteren Flächen enthalten, nämlich  $f_{22}$  und  $f_{32}$ . Die Grenzen dieser Flächen verlaufen durch  $v_u$ , und die Flächen selbst gehören zur Menge  $L(v_u)$ . Wie bestimmen wir nun  $L(v_u)$ ?

Rekapitulieren wir, was wir über die Flächen in  $L(v_u)$  wissen: Zunächst soll jede Fläche in  $L(v_u)$  den aktuellen Eventpunkt  $v$  ‘in ihrem Innern’ oder aber auf ihrer Begrenzung enthalten; das ist nach Hilfssatz 3.3 ‘automatisch’ der Fall. Zweitens soll die Begrenzung jeder Fläche in  $L(v_u)$  durch  $v_u$  verlaufen, alle in Frage kommenden Flächengrenzen sind also an  $v_u$  gebunden und uns daher bekannt. Drittens muß die Begrenzung jeder Fläche in  $L(v_u)$  um  $v$  gegen den Uhrzeigersinn orientiert sein. Wir müssen also nur für jeden der ‘ursprünglichen’ Halbkantenzüge, der durch  $v_u$  verläuft, die Orientierung ermitteln; ist ein Halbkantenzug gegen den Uhrzeigersinn orientiert, verweisen die *innereFläche()*-Zeiger auf genau die Flächen, die zu  $L(v_u)$  gehören.

Da die *Vorgänger()*- und *Nachfolger()*-Zeiger aller Halbkanten um  $v_u$  bereits aktualisiert sind, können wir die ‘ursprünglichen’ Halbkantenzüge jedoch nicht mehr rekonstruieren; selbst wenn wir das könnten, würde die Bestimmung der Orientierung viel zu viel Zeit kosten (erinnern wir uns, daß wir, um die Orientierung eines Halbkantenzuges zu ermitteln, zunächst den am weitesten links liegenden Knoten suchen müssen! vgl. S. 68). Die Frage ist also, wie die Orientierung der durch  $v_u$  verlaufenden Flächengrenzen allein aufgrund der Lage von  $v$  und  $v_u$  festgestellt werden kann.

Die Lösung veranschaulicht Abb. 3.16 (rechts).<sup>22</sup> Sämtliche Kanten, die an  $v_u$  gebunden sind, sind dort in Form ihrer Repräsentation durch die doppelt-verkettete Kantenliste  $\mathcal{D}$  dargestellt, und zwar zum Zeitpunkt *nach* der Aktualisierung der Knoten- und Kantenliste und *vor* der Aktualisierung der Flächeninformation. Wir können zunächst feststellen, daß die *innereFläche()*-Zeiger von je zwei Halbkanten auf ein und dieselbe Fläche zeigen. Es ist klar, daß diese Beobachtung an *jedem* Knoten aus  $Sub(\mathcal{S}_{1..m})$  gilt, unabhängig davon, ob es sich um einen Schnittpunkt handelt, oder, wie in diesem Fall, um einen Knoten, der bereits in  $\bigcup_{i=1}^m \mathcal{S}_i$  existiert. Jedes dieser Halbkantenpaare repräsentiert einen Teil der Grenze von je einer Fläche aus  $\bigcup_{i=1}^m \mathcal{S}_i$ . Eine Halbkante jedes solchen Paares hat  $v_u$  zum Ziel, bei der anderen liegt  $v_u$  im Ursprung. Wir wissen, daß jeder ‘ursprüngliche’ Halbkantenzug, der eine Fläche aus  $\bigcup_{i=1}^m \mathcal{S}_i$  begrenzt, die  $v$  enthält, gegen den Uhrzeigersinn orientiert sein muß.

<sup>22</sup>Beim rechten Teil von Abb. 3.16 handelt es sich um einen zweifach vergrößerten Ausschnitt der Flächengrenzen im linken Teil der Abbildung.

Wenn wir nun die Perspektive des Knotens  $v$  einnehmen, bedeutet das, daß dieser Halbkantenzug von ‘rechts’ kommend über  $v_u$  nach ‘links’ laufen muß. Wenn wir also von  $v$  aus den Knoten  $v_u$  gegen den Uhrzeigersinn umlaufen, müßten wir zuerst auf diejenige der beiden Halbkanten treffen, die  $v_u$  zum Ziel hat. Umgekehrt treffen wir von  $v$  aus im Uhrzeigersinn zuerst auf diejenige der beiden Halbkanten, die  $v_u$  ‘verläßt’ (deren Ursprung also  $v_u$  ist). Bei einem Halbkantenzug, der (von  $v$  aus gesehen) im Uhrzeigersinn orientiert ist, finden wir die ‘ein-Ć und ausgehenden’ Halbkanten genau in umgekehrter Reihenfolge.

Diese Überlegung führt uns direkt zur Lösung des Problems. Wir umlaufen, mit  $v$  beginnend,  $v_u$  im Uhrzeigersinn und führen dabei zwei Listen, die wir mit  $O(v_u)$  und  $T(v_u)$  bezeichnen. Jedesmal, wenn wir auf eine Kante stoßen, schreiben wir die innere Fläche derjenigen Halbkante, deren Ursprung in  $v_u$  liegt, in die Liste  $O(v_u)$ , und die innere Fläche derjenigen Halbkante, deren Ziel  $v_u$  ist, in die Liste  $T(v_u)$ . Wenn wir  $v_u$  vollständig umlaufen haben, enthalten beide Listen die gleiche Anzahl von Einträgen, und zwar jeweils exakt so viele, wie Kanten an  $v_u$  gebunden sind. Unserer obigen Argumentation folgend muß nun jede Fläche, deren Begrenzung von  $v$  aus gesehen gegen den Uhrzeigersinn orientiert ist, zuerst in die Liste  $O(v_u)$  und erst danach in die Liste  $T(v_u)$  eingetragen worden sein. Mit anderen Worten: Alle die Flächen, deren Begrenzungen durch  $v_u$  verlaufen und um  $v$  gegen den Uhrzeigersinn orientiert sind, stehen in  $O(v_u)$  auf einem niedrigeren Rang als in  $T(v_u)$ .

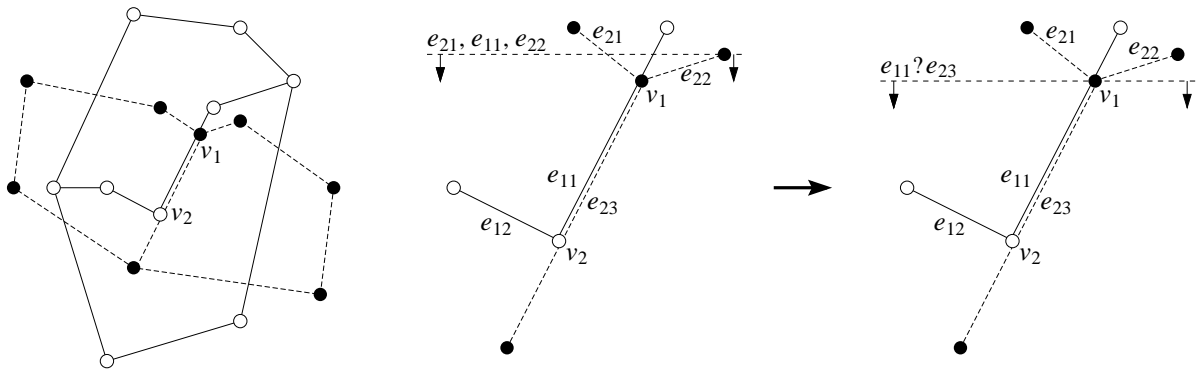
In Abb. 3.16 (rechts) stehen die Flächen  $f_{41}$ ,  $f_{32}$  und  $f_{22}$  in  $O(v_u)$  auf einem niedrigeren Rang als in  $T(v_u)$ . Damit ist die gesuchte Menge  $L(v_u) = \{f_{41}, f_{32}, f_{22}\}$ . Wir kennen bereits  $E(v_u) = \{f_{11}\}$  und  $B(v) = \{f_{40}, f_{41}\}$  und können die Flächen aus  $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$ , die den Knoten  $v$  enthalten, ermitteln als  $E(v) = \{E(v_u) \cup L(v_u)\} \setminus B(v) = \{f_{11}, f_{32}, f_{22}\}$ .

Wenn wir die Menge  $L(v_u)$  bestimmen, indem wir alle Flächen in  $O(v_u)$  suchen, die dort auf einer niedrigeren Position stehen als in  $T(v_u)$ , benötigen wir allerdings quadratische Zeit: Wenn wir den Grad des Knotens  $v_u$  mit  $n_u$  bezeichnen, benötigt die Bestimmung der Menge  $L(v_u)$  nach dem oben beschriebenen Verfahren  $n_u^2$  Schritte. Die Idee der zwei Listen diente jedoch ausschließlich didaktischen Zwecken, denn der Praxis implementieren wir sowohl  $O(v_u)$  als auch  $T(v_u)$  als balancierten Binärbaum und fügen nur dann eine Fläche in  $O(v_u)$  ein, wenn diese Fläche noch nicht in  $T(v_u)$  eingetragen wurde. Wenn wir  $v_u$  vollständig umlaufen haben, enthält  $O(v_u)$  genau die Flächen, die zur Menge  $L(v_u)$  gehören. Auf diese Weise liegt die Zeit, die zur Bestimmung der Menge  $L(v_u)$  benötigt wird, in  $O(n_u \log n_u)$ . Die Menge  $E(v_u)$  ist bekannt, die Bestimmung der Menge  $B(v)$  erfordert eine Zeit, die linear zum Grad von  $v$  ist. Daher benötigt die Bestimmung der einen Knoten  $v \in \text{Sub}(\mathcal{S}_{1\dots m})$  umgebenden Flächen eine Zeit in  $O(n_u \log n_u)$ .

### 3.3.4 Zeitaufwand

Bis auf die Aktualisierung der Flächeninformation entspricht das Verfahren zur Verschneidung von  $m$  Flächenkarten dem Verfahren von nur zwei Flächenkarten. Berücksichtigt man den Zeitaufwand zur Aktualisierung der Flächenlabel nicht, liegt die zur Verschneidung von  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  benötigte Zeit entsprechend in  $O(n \log n + k \log n)$ , wobei  $n$  die Summe der Komplexitäten von  $\mathcal{S}_{1\dots m}$  ist und  $k$  die Komplexität der Verschneidung (vgl. Satz 3.2 auf S. 73). Die zur Aktualisierung der Flächeninformationen zusätzlich benötigte Zeit liegt in  $O(k \log k)$ .

**Satz 3.5.** Seien  $\mathcal{S}_{1\dots m}$   $m$  planare Subdivisionen der Komplexitäten  $n_1, n_2, \dots, n_m$ , und sei  $n := \sum_{i=1}^m n_i$ . Wenn sich keine zwei Kanten aus  $\bigcup_{i=1}^m \mathcal{S}_i$  überlappen, dann liegt die zur Berechnung der Verschneidung von  $\bigcup_{i=1}^m \mathcal{S}_i$  benötigte Zeit in  $O(n \log n + k \log n + k \log k)$ , wobei  $k$  die Komplexität der Verschneidung ist.



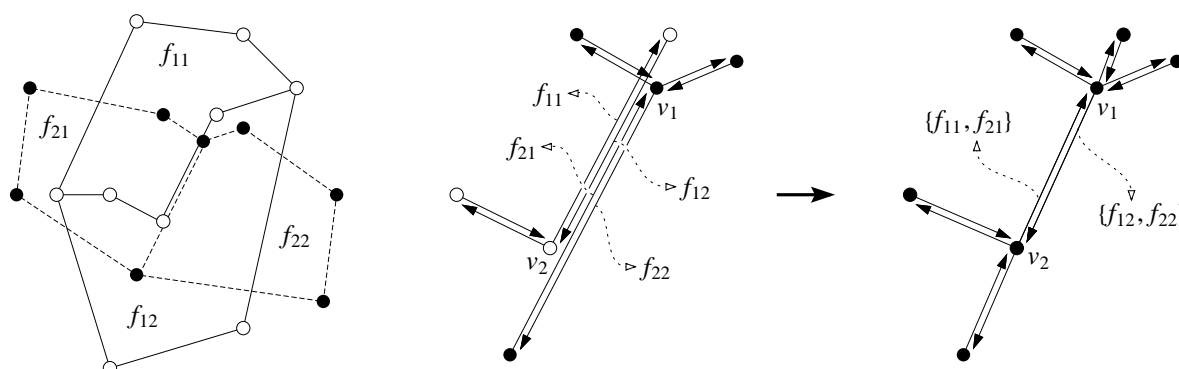
**Abbildung 3.17** Das Problem zweier sich überlappender Kanten. Dargestellt sind zwei überlagerte Subdivisionen  $\mathcal{S}_1$  (gerissene Kanten mit schwarzen Knoten) und  $\mathcal{S}_2$  (durchgezogene Kanten mit weißen Knoten). Paare von Kanten, die sich überlappen (hier:  $e_{11}$  und  $e_{23}$ ), können beim Scanline-Verfahren auf einfache Weise ermittelt werden, da sie in der Statusliste  $\mathcal{T}$  nicht unterscheidbar sind.

### 3.4 Die Handhabung sich überlappender Kanten

Bislang haben wir ein Problem außer acht gelassen, das in der Praxis relativ häufig auftritt und für unsere Erörterungen in Kap. 3.5 von entscheidender Bedeutung sein wird: Das Problem sich überlappender Kanten (Abb. 3.18 links; vgl. auch Edelsbrunner, Mücke 1990; Schorn 1994; Emiris, Canny 1995). Die Überlappung zweier Kanten bedeutet nichts weiter, als daß sich die Kanten in mehr als einem Punkt schneiden. Eine Überlappung stellt somit einen degenerierten Überschneidungsfall dar, weshalb die Kenntnis aller sich überlappenden Kanten für die Bestimmung des der planaren Subdivision  $Sub(\mathcal{S}_1 \dots \mathcal{S}_m)$  zugrundeliegenden Graphen von ähnlich grundlegender Bedeutung ist wie die Kenntnis aller Schnittpunkte. Folgen wir dem Scanline-Prinzip (Kap. 3.2.1), wird jeder Schnittpunkt, der (der Ordnung  $\prec_Q$  entsprechend) ‘unterhalb’ des aktuellen Eventpunktes gefunden wird, als neuer Eventpunkt in die Eventfolge  $Q$  eingefügt. Wird nun eine Überlappung gefunden, kann diesem Prinzip nicht unmittelbar Folge geleistet werden, da jede Überlappung ja aus ‘unendlich vielen’ Punkten besteht. Das Scanline-Prinzip erfordert zudem die ständige Aktualisierung der Statusfolge  $\mathcal{T}$ , die die lokale Ordnung der Kanten auf bzw. direkt unterhalb einer imaginären, den aktuellen Eventpunkt schneidenden Scanlinie repräsentiert. In welcher Reihenfolge stehen aber zwei sich überlappende Kanten auf der Scanlinie? Hier liegen die entscheidenden Schwierigkeiten, doch zugleich liegt hier auch der Schlüssel zur Lösung des Problems.

#### 3.4.1 Bestimmung sich überlappender Kantenpaare

Betrachten wir Abb. 3.17. Dargestellt sind zwei überlagerte planare Subdivisionen  $\mathcal{S}_1$  und  $\mathcal{S}_2$  (links); die Kante  $e_{11} \in \mathcal{S}_1$  überlappt sich partiell mit der Kante aus  $e_{23} \in \mathcal{S}_2$ . Die Überlappung ‘beginnt’ im Knoten  $v_1 \in \mathcal{S}_2$  und ‘endet’ im Knoten  $v_2 \in \mathcal{S}_1$ . Am Eventpunkt direkt vor  $v_1$  ist die Ordnung der Kanten in der Statusfolge  $\mathcal{T}$  eindeutig (Mitte): Von links nach rechts wird die Scanlinie von den Kanten  $e_{21}$ ,  $e_{11}$  und  $e_{22}$  geschnitten. Wird nun als nächster Eventpunkt  $v_1$  erreicht, werden alle Kanten, deren unterer Endpunkt auf der neuen Scanlinie liegt, aus  $\mathcal{T}$  entfernt (rechts; vgl. BEARBEITEEVENTPUNKT auf S. 60). In  $\mathcal{T}$  verbleibt also die Kante  $e_{11}$ . Anschließend werden alle Kanten, deren oberer Endpunkt im aktuellen Eventpunkt liegt, in  $\mathcal{T}$  eingefügt. Im dargestellten Fall soll also die Kante  $e_{23}$  zusätzlich zur Kante  $e_{11}$  in  $\mathcal{T}$  eingefügt werden. Nun unterscheiden sich  $e_{11}$  und  $e_{23}$  auf der Scanlinie nicht, da sie diese im selben Punkt  $v_1$  schneiden. Aber auch unterhalb von  $v_1$  lassen sich die beiden Kanten

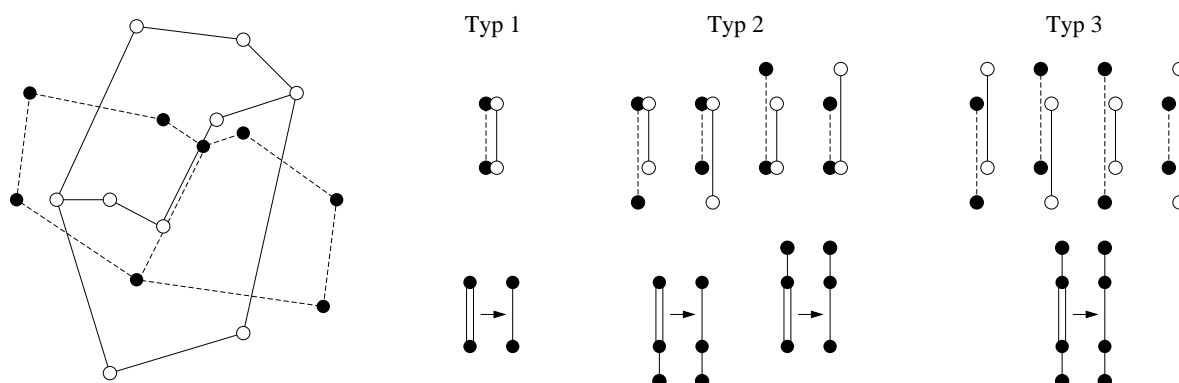


**Abbildung 3.18** Die Aktualisierung von überlappenden Kantenpaaren. Wenn zwei planare Subdivisionen ein sich überlappendes Kantenpaar aufweisen (links), ist der Überlappungsbereich in der doppelt-verketteten Kantenliste  $\mathcal{D}$  zunächst durch vier Halbseiten repräsentiert, von denen jede auf genau eine innere Fläche zeigt (Mitte). Nach der Aktualisierung von  $\mathcal{D}$  existieren nur noch zwei Halbseiten, deren *innereFläche()*-Zeiger nun auf Mengen von Flächen verweisen (rechts).

nicht im Sinne von ‘links von’ und ‘rechts von’ anordnen, da sie sich ja überlappen. Mit anderen Worten: Die überlappenden Kanten  $e_{11}$  und  $e_{23}$  sind am aktuellen Eventpunkt in  $\mathcal{T}$  nicht unterscheidbar. Umgekehrt bedeutet das, daß wir Paare sich überlappender Kanten auf einfache Weise finden können, indem wir prüfen, ob eine Kante, die in  $\mathcal{T}$  eingefügt werden soll, bereits in  $\mathcal{T}$  ‘vorhanden’ ist: Da wir die Statusfolge  $\mathcal{T}$  als balancierten Binärbaum implementiert haben, können wir nach  $e_{23}$  suchen. Als Suchergebnis erhalten wir  $e_{11}$ . Nun ist  $e_{11} \neq e_{23}$ , und wir können schließen, daß beide Kanten nicht unterscheidbar sind und sich daher überlappen.

### 3.4.2 Aktualisierung der Flächenlabel

Da wir im Verlauf des Scanline-Verfahrens der Eventfolge  $\mathcal{Q}$  folgen, finden wir immer zuerst den ‘oberen Anfangspunkt’ eines Überlappungsbereiches. Wir müssen daher nur darum Sorge tragen, daß wir die gefundene Überlappung unterhalb des aktuellen Eventpunktes auflösen. Es ist klar, daß in  $Sub(\mathcal{S}_{1\dots m})$  jeder gefundene Überlappungsbereich durch genau eine Kante repräsentiert werden muß. Was daher zu geschehen hat, verdeutlicht Abb. 3.18. Dargestellt ist derselbe Fall wie in Abb. 3.17 (links), nur in Form seiner Repräsentation durch die doppelt-verkettete Kantenliste  $\mathcal{D}$  (Mitte und rechts). Es überlappen sich zwei Kanten aus  $\mathcal{S}_1 \cup \mathcal{S}_2$ , von denen jede genau zwei Flächen begrenzt; der Überlappungsbereich ist in  $\mathcal{D}$  also durch vier Halbseiten repräsentiert, deren *innereFläche()*-Zeiger auf jeweils genau eine Fläche verweisen (Mitte). Nach Auflösung der Überlappung wird der ehemalige Überlappungsbereich nur noch durch zwei Halbseiten repräsentiert, die genau einer Kante aus  $Sub(\mathcal{S}_{1\dots m})$  entsprechen (rechts). Auf beiden Seiten dieser Kante liegen genau zwei ‘Originalflächen’, so daß nun jeder der *innereFläche()*-Zeiger der beiden Halbseiten auf zwei Flächen verweisen muß. Bei der Verschneidung von  $m$  planaren Subdivisionen können natürlich auch Fälle auftreten, in denen sich eine unbestimmte Zahl  $n_l$  ( $2 \leq n_l \leq m$ ) von Kanten überlappt. In solchen Fällen müssen die *innereFläche()*-Zeiger auf  $n_l$  ‘Originalflächen’ verweisen. Im Falle von Überlappungen müssen wir daher unsere Vorstellung der *innereFläche()*-Zeiger revidieren: Sie verweisen nun nicht mehr auf einzelne Flächen, sondern auf Mengen von Flächen.



**Abbildung 3.19** Eine Typisierung von Überlappungsproblemen. Bei paarweisen Überlappungen können drei Problemtypen unterschieden werden (obere Reihe). Jedes Überlappungsproblem kann durch geeignete Teilung der Kanten in ein Problem des Typs 1 transformiert werden (untere Reihe). Bei unserem Beispiel (links) handelt es sich um ein Problem des Typs 3.

### 3.4.3 Eine Typisierung von Überlappungsproblemen

Die Anzahl  $n_l$  der sich überlappenden Kanten ist für den Umgang mit Überlappungsfällen zum Glück völlig irrelevant, vorausgesetzt, wir reagieren beim Auftreten einer paarweisen Überlappung unmittelbar: Wir können nur für jeweils eine Kante prüfen, ob eine überlappende Kante bereits in  $\mathcal{T}$  vorhanden ist; wir finden also entweder eine paarweise oder gar keine Überlappung. Wenn wir eine paarweise Überlappung sofort auflösen, erhalten wir als Ergebnis eine einzige Kante. Entsprechend führt die Überlappung dieser neuen mit einer weiteren Kante wiederum zu einer einzigen neuen Kante. Bei dieser Vorgehensweise werden wir also im Laufe des Verfahrens ausschließlich mit paarweisen Überlappungen konfrontiert, deren Auflösung wiederum jeweils zwei Operationen erfordern: Die überlappenden Kanten müssen geteilt und die *innereFläche()*-Zeiger aktualisiert werden. Die Aktualisierung der inneren Flächen ist natürlich nur für den Überlappungsbereich selbst erforderlich. Wir können das Problem daher noch einfacher strukturieren, indem wir sich überlappende und sich nicht überlappende Bereiche getrennt bearbeiten. In Abb. 3.19 (obere Reihe) ist leicht zu sehen, daß sich die möglichen Überlappungen zweier verschiedener Kanten in drei verschiedene Typen untergliedern lassen:

1. Die beiden Kanten haben genau zwei gemeinsame Endpunkte, d.h. die Kanten überlappen sich vollständig, und eine Teilung von Kanten ist nicht erforderlich.
2. Die beiden Kanten haben genau einen gemeinsamen Endpunkt, so daß die längere der beiden Kanten in einem Knoten geteilt werden muß.
3. Die beiden Kanten haben keinen gemeinsamen Endpunkt, so daß entweder jede der beiden Kanten in je einem Knoten oder eine der beiden Kanten in zwei Knoten geteilt werden muß.

Überlappungsprobleme des Typs 1 führen zu genau einer neuen Kante in  $Sub(\mathcal{S}_{1\dots m})$ , Probleme des Typs 2 zu genau zwei und Probleme des Typs 3 zu genau drei neuen Kanten in  $Sub(\mathcal{S}_{1\dots m})$  (Abb. 3.19 untere Reihe; bei unserem Beispiel links handelt es sich also um ein Problem des Typs 3). Führen wir zunächst die notwendigen Teilungen wie in Kap. 3.2.3.1 besprochen durch, ohne die Überlappung aufzulösen, erhalten wir immer ein Kantenpaar, das sich vollständig überlappt und damit dem Problemtyp 1 entspricht, und je nach vorliegendem Problemtyp eine Anzahl von Kanten, die sich nicht überlappen (zwei beim Typ 3, eine beim Typ 2 und keine beim Typ 1). Alle Überlappungsprobleme können so in ein Problem des Typs 1 transformiert werden.

### 3.4.4 Die algorithmische Bearbeitung von Überlappungsproblemen

Fassen wir zusammen: Während des Scanline-Verfahrens können zwei sich überlappende Kanten gefunden werden, wenn beim Einfügen einer Kante  $e$  in die Statusfolge  $\mathcal{T}$  geprüft wird, ob diese Kante in  $\mathcal{T}$  von allen anderen Kanten aus  $\mathcal{T}$  unterscheidbar ist. Ist eine Kante in  $\mathcal{T}$  nicht unterscheidbar, kennen wir die Kante  $e_{\mathcal{T}}$ , die sich mit  $e$  überlappt. Um die Überlappung aufzulösen, entfernen wir zunächst die Kante  $e_{\mathcal{T}}$  von der Statusfolge und die beiden Kanten  $e$  und  $e_{\mathcal{T}}$  von den Eventpunkten, die in den oberen Endpunkten der beiden Kanten liegen. Anschließend teilen wir beide Kanten je nach Überlappungstyp in den entsprechenden Knoten und erhalten eine Menge von maximal vier neuen Kanten, von denen sich genau zwei vollständig überlappen. Die Kanten, die sich in dieser neuen Menge nicht überlappen, speichern wir mit den Eventpunkten, die in den oberen Endpunkten dieser Kanten liegen. Das sich überlappende Kantenpaar vereinigen wir schließlich zu einer neuen Kante, die alle Flächen voneinander trennt, die von beiden Kanten dieses Paares getrennt werden.

Von einer algorithmischen Umsetzung der so skizzierten Vorgehensweise erwarten wir, daß sie sich in das Scanline-Prinzip (das ja gewissermaßen das ‘Herzstück’ unseres Verfahrens bildet) integrieren läßt, ohne es grundlegend zu verändern. Die Umsetzung beginnt daher mit einer Modifikation der Zeile 7 in BEARBEITEEVENTPUNKT (vgl. S. 60). Statt alle Kanten, deren obere Endpunkte im aktuellen Eventpunkt liegen oder die den aktuellen Eventpunkt enthalten, direkt in die Statusfolge  $\mathcal{T}$  einzufügen, gehen wir nun den ‘Umweg’ über eine externe Prozedur:

**Algorithmus** KANTEINSTATUSEINFÜGEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* keine.

1. **if** ( $e$  ist im aktuellen Eventpunkt in der Statusfolge  $\mathcal{T}$  gültig)  
**then**
2.     Suche nach  $e$  in der Statusfolge  $\mathcal{T}$ . Sei  $e_{\mathcal{T}}$  die Kante, die bei der Suche gefunden wird.
3.     **if** ( $e_{\mathcal{T}}$  existiert und  $e_{\mathcal{T}} \neq e$ )  
      **then**
4.         PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e, e_{\mathcal{T}}$ ).  
      **else**
5.         Füge  $e$  in  $\mathcal{T}$  ein.  
      **end if**
- end if**

Von ‘außen’ gesehen muß jeder Aufruf von KANTEINSTATUSEINFÜGEN genauso wirken, als würden wir eine Kante direkt in  $\mathcal{T}$  einfügen. Zunächst wird kontrolliert, ob die Kante am aktuellen Eventpunkt in  $\mathcal{T}$  gültig ist. Diese Kontrolle ist im Moment ohne Belang, wird jedoch notwendig, wenn wir in Kap. 3.5 mit diskreten Koordinaten arbeiten. Eine Kante  $e = (v_u, v_l)$  ist natürlich genau dann am aktuellen Eventpunkt  $v$  in  $\mathcal{T}$  gültig, wenn  $v_u$  der obere Endpunkt von  $e$  ist und  $v_u \preceq_Q v \wedge v \prec_Q v_l$ . Wenn keine Überlappung vorliegt, läuft das Verfahren auch weiter wie bislang besprochen. Der Überlappungstest beim Einfügen einer Kante in  $\mathcal{T}$  erhöht zunächst die Laufzeit des Verfahrens nicht, denn neben der ohnehin notwendigen Einfügeoperation wird genau eine Suchoperation in  $\mathcal{T}$  ausgeführt. Im Falle eines positiven Überlappungstests muß das sich überlappende Kantenpaar jedoch gesondert bearbeitet und erst das Ergebnis dieser ‘Sonderbehandlung’ in  $\mathcal{T}$  eingefügt werden:

**Algorithmus** PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e_1, e_2$ )

*Eingabe:* Zwei sich überlappende Kanten  $e_1$  und  $e_2$ .

*Ausgabe:* keine.

1. Lösche  $e_1$  und  $e_2$  aus  $\mathcal{T}$  und aus  $\mathcal{D}$ .
2. KANTEAUSEVENTFOLGELÖSCHEN( $e_1$ )

3. KANTEAUSEVENTFOLGELÖSCHEN( $e_2$ )
4. Teile die Kanten  $e_1$  und  $e_2$  je nach Überlappungstyp an den entsprechenden Knoten, wie oben beschrieben. Sei  $E$  die Menge der Kanten, die sich durch die Teilung ergeben, seien  $e'_1 \in E$  und  $e'_2 \in E$  die beiden neuen Kanten, die sich jetzt vollständig überlappen, und sei  $K = E \setminus \{e'_1, e'_2\}$  die Teilmenge der neuen Kanten, die sich nicht überlappen. ( $K$  enthält keine, eine oder zwei Kanten.)
5. **for each** ( $e_i \in K$ ) **do**
6.     KANTEINEVENTFOLGEINFÜGEN( $e_i$ )
7.     KANTEINSTATUSFOLGEINFÜGEN( $e_i$ )
8. **end for**
8. VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN( $e'_1, e'_2$ )

Die beiden beteiligten Kanten werden aus der Eventfolge  $\mathcal{Q}$  und aus der Statusfolge  $\mathcal{T}$  entfernt – letzteres unabhängig davon, ob sie in  $\mathcal{T}$  existieren. Anschließend wird das Überlappungsproblem auf ein Problem des Typs 1 reduziert und alle ‘Reststücke’ in der Eventfolge aktualisiert. Mit Blick auf die Diskussion in Kap. 3.5 wurden hier bereits die Operationen, die auf der Eventfolge  $\mathcal{Q}$  durchgeführt werden, als eigenständige Prozeduren ausgegliedert. Ihre Aufgabe ist im Moment trivial und besteht nur darin, diejenigen Kanten, die mit den Eventpunkten gespeichert sind, zu löschen bzw. einzufügen:

**Algorithmus** KANTEAUSEVENTFOLGELÖSCHEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* keine.

1. Sei  $v$  der obere Endpunkt der Kante  $e$ .
2. Suche  $v$  in der Eventfolge  $\mathcal{Q}$ .
3. Lösche  $e$  aus der Menge der Kanten, die mit  $v$  gespeichert sind.

**Algorithmus** KANTEINEVENTFOLGEINFÜGEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* keine.

1. Sei  $v$  der obere Endpunkt der Kante  $e$ .
2. Suche  $v$  in der Eventfolge  $\mathcal{Q}$ .
3. Speichere  $e$  mit  $v$ .

Die beiden Kanten, die sich nun vollständig überlappen, müssen schließlich zu einer Kante ‘vereinigt’ werden. Diese Kante verweist auf alle ursprünglichen Flächeninformationen und muß in der Eventfolge  $\mathcal{Q}$  aktualisiert und in die Statusliste  $\mathcal{T}$  eingefügt werden (denn der aktuelle Eventpunkt ist immer der obere Endpunkt der beiden sich vollständig überlappenden Kanten!):

**Algorithmus** VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN( $e_1, e_2$ )

*Eingabe:* Zwei Kanten  $e_1$  und  $e_2$ , die sich vollständig überlappen.

*Ausgabe:* keine.

1. Sei  $v_1$  der obere Endpunkt von  $e_1$  und  $e_2$ , sei  $v_2$  der untere Endpunkt von  $e_1$  und  $e_2$ . Seien  $h_{1d}$  und  $h_{2d}$  die beiden Halbkanten von  $e_1$  und  $e_2$ , deren Ursprung in  $v_1$  liegt, und seien  $h_{1u}$  und  $h_{2u}$  die beiden Halbkanten von  $e_1$  und  $e_2$ , deren Ursprung in  $v_2$  liegt. Seien schließlich  $F_{1d}$ ,  $F_{2d}$ ,  $F_{1u}$  und  $F_{2u}$  die Mengen der Flächen, auf die die *innereFläche()*-Zeiger von  $h_{1d}$ ,  $h_{2d}$ ,  $h_{1u}$  und  $h_{2u}$  verweisen.
2. Sei  $e = (v_1, v_2)$  eine neue Kante. Sei  $h_d$  die Halbkante von  $e$ , deren Ursprung in  $v_1$  liegt, und sei  $h_u$  die Halbkante von  $e$ , deren Ursprung in  $v_2$  liegt.
3. Setze die *innereFläche()*-Zeiger der Halbkanten von  $e$ , so daß *innereFläche*( $h_d$ ) =  $F_{1d} \cup F_{2d}$  und *innereFläche*( $h_u$ ) =  $F_{1u} \cup F_{2u}$ .
4. KANTEINSTATUSEINFÜGEN( $e$ )
5. KANTEINEVENTFOLGEINFÜGEN( $e$ )

Wir sehen hier, daß die beiden Prozeduren KANTEINSTAUSEINFÜGEN und VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN im Prinzip ein rekursives Funktionenpaar darstellen: Unter gewissen Umständen können sich die Prozeduren mehrmals wechselseitig aufrufen. Diese Form der Implementierung erfolgt an dieser Stelle ebenfalls einzig mit Blick auf die Rundungsproblematik, die im folgenden Kapitel besprochen wird. Für den Moment wird eine Rekursionstiefe von 1 niemals überschritten, denn für den Fall, daß eine Überlappung in KANTEINSTAUSEINFÜGEN entdeckt wird, erfolgt nur genau ein weiterer Aufruf von KANTEINSTAUSEINFÜGEN. Bei diesem zweiten (dem ersten rekursiven) Aufruf in Zeile 4 in VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN fällt der Test auf Überlappung in jedem Fall negativ aus, denn die beiden überlappenden Kanten wurden zuvor aus der Statusfolge  $\mathcal{T}$  entfernt, und sollten sich noch weitere Kanten im und unterhalb des aktuellen Eventpunktes überlappen, wurde noch nicht versucht, diese in  $\mathcal{T}$  einzufügen. Beim augenblicklichen Stand unserer Diskussion könnte also Zeile 4 in VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN auch lauten ‘Füge  $e$  in  $\mathcal{T}$  ein’.

### 3.4.5 Zeitaufwand

Fragen wir uns abschließend, welchen zusätzlichen Zeitaufwand die Auflösung von Überlappungsproblemen erfordert. Wir können zunächst feststellen, daß die Handhabung von Überlappungsproblemen gewissermaßen ‘Problem-sensitiv’ ist, denn wir führen nur dann zusätzliche Operationen aus, wenn tatsächlich eine Überlappung vorliegt (wie oben bereits erwähnt, beeinflußt die zusätzliche Suche in  $\mathcal{T}$  die Gesamtlaufzeit nicht). Jeder Überlappungsfall erfordert zunächst jeweils zwei Löschoptionen in  $\mathcal{Q}$  und in  $\mathcal{T}$  (Zeilen 1-3 in PARTIELLEÜBERLAPPUNGAUFLÖSEN) sowie maximal drei Einfügeoperationen in  $\mathcal{Q}$  (Zeile 7 in PARTIELLEÜBERLAPPUNGAUFLÖSEN und Zeile 5 in VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN). Jede dieser Operationen erfordert  $O(\log k)$  Zeit, wobei  $k$  die Komplexität der Verschneidung  $Sub(\mathcal{S}_{1\dots m})$  ist. Die zur Bestimmung des Überlappungstyps benötigte Zeit ist konstant (Zeile 4 in PARTIELLEÜBERLAPPUNGAUFLÖSEN; es müssen genau vier Punkte sortiert und maximal vier neue Kanten erzeugt werden). Die Zeit, die benötigt wird, um die Vereinigungsmengen der inneren Flächen in Zeile 3 in VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN zu bestimmen, ist abhängig von der Anzahl  $n_l$  der Kanten, die sich in einem Eventpunkt überlappen und liegt in  $O(n_l \log n_l)$ . Daher gilt der folgende Satz (vgl. Satz 3.3.4 auf S. 82):

**Satz 3.6.** *Seien  $\mathcal{S}_{1\dots m}$   $m$  planare Subdivisionen der Komplexitäten  $n_1, n_2, \dots, n_m$ , und sei  $n := \sum_{i=1}^m n_i$ . Die zur Berechnung der Verschneidung von  $\bigcup_{i=1}^m \mathcal{S}_i$  benötigte Zeit liegt in  $O(n \log n + k \log n + k \log k + l \log k)$ , wobei  $k$  die Komplexität der Verschneidung und  $l$  die Anzahl der sich überlappenden Kanten ist.*

## 3.5 Umgang mit unpräziser Arithmetik

Aufgrund der bisherigen Ausführungen in diesem Kapitel dürfen wir uns nun im Prinzip in die Lage versetzt fühlen, jedes allgemeine Flächenverschneidungsproblem lösen zu können: Wir können beliebig viele Flächenkarten ‘übereinanderlegen’ und diese verschneiden, wobei wir auch degenerierte Fälle, wie kollineare Knoten, senkrechte, waagerechte oder gar sich überlappende Kanten usw., im Griff haben. Bei unserer Analyse des Problems wurde zu Beginn versucht, solche Ausnahmefälle zu ignorieren; wir haben das Problem zunächst auf ein Segmentschnittproblem reduziert und konnten so den Scanline-Algorithmus als Lösungsverfahren einführen. Dieser wurde dann für unsere Zwecke zu einem Verfahren der zweifachen Flächenverschneidung ausgebaut (Kap. 3.2). In einem nächsten Schritt wurde das gewonnene Verfahren auf die Verschneidung beliebig vieler Flächenkarten verallgemeinert (Kap. 3.3). Wir haben gesehen, daß das allgemeine Flächenverschneidungsproblem vom geometrischen Aspekt her identisch ist mit dem klassischen (zweifachen) Verschneidungsproblem;



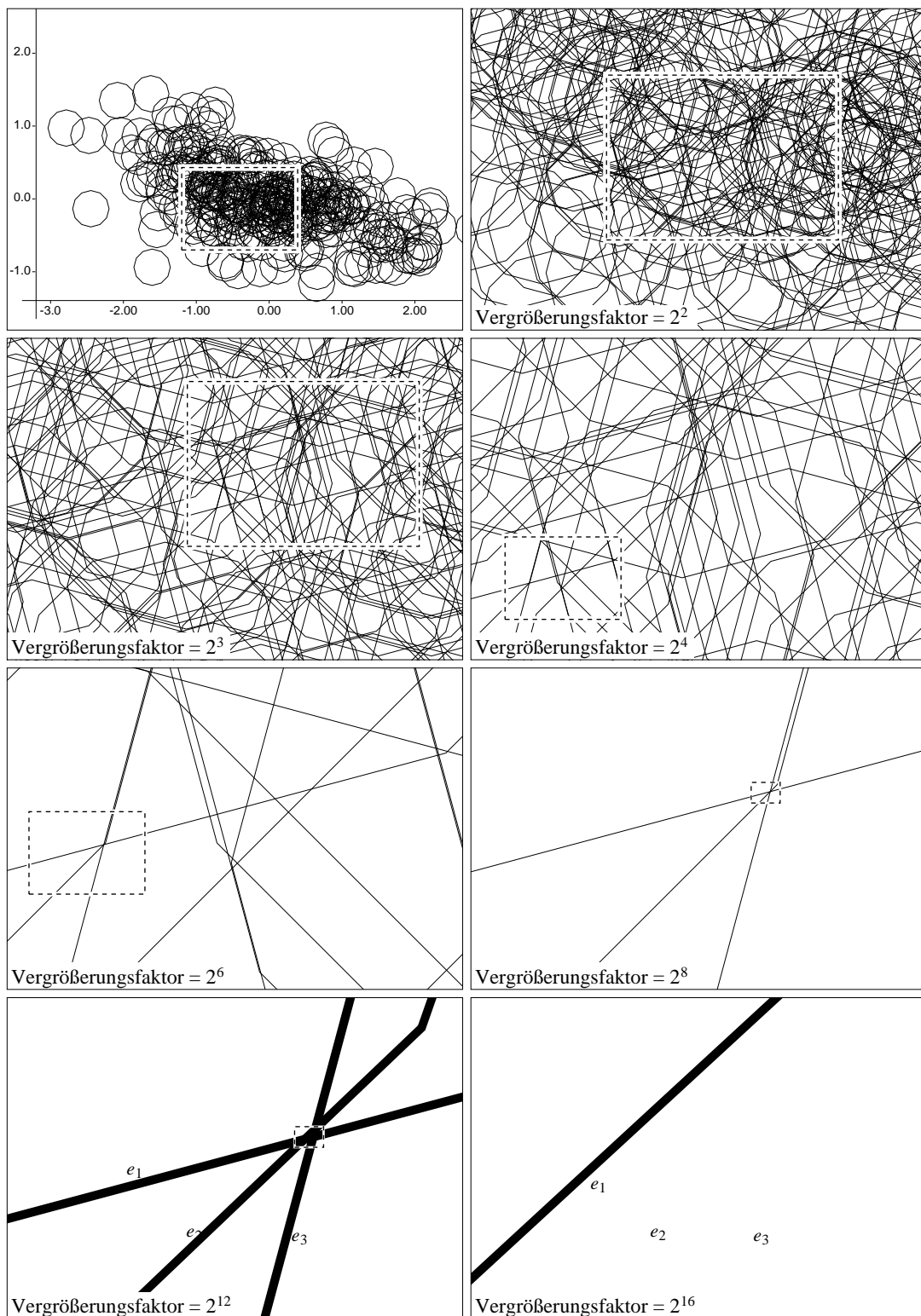
einzig die Aktualisierung der Flächenlabel erforderte entsprechenden Mehraufwand, und mit der Lösung des Problems der umgebenden Flächen haben wir das allgemeine Flächenverschneidungsproblem auf *symbolischer* Ebene schließlich *konzeptionell* vollständig durchdrungen.

Die Phase der konzeptionellen Klärung des Problems und seiner Lösung ist notwendig, jedoch in der Regel nicht hinreichend für eine praktische Implementierung des Verfahrens. In der Praxis treten häufig sogenannte *degenerierte Fälle* auf, die ein Problem komplizieren und zusätzliche Operationen auf dem Weg zur Problemlösung erfordern. In unserem Fall stellen sich überlappende Kanten solche degenerierten Fälle dar, und wir haben gesehen, welche zusätzlichen *symbolischen* Operationen ihre Existenz erfordert (Kap. 3.4). Weitere degenerierte Fälle sind für das allgemeine Flächenverschneidungsproblem nicht auszumachen, und eigentlich könnten wir nun entspannt damit beginnen, das beschriebene Lösungsverfahren in einer beliebigen Programmiersprache konkret zu implementieren.

Leider steckt der Teufel jedoch wie immer im Detail. Die Implementation des Verfahrens, so wie es bislang vorgestellt wurde, führt zu einer Software, mit der in der Tat jedes allgemeine Flächenverschneidungsproblem gelöst werden kann, vorausgesetzt, die zur Problemlösung erforderliche Rechengenauigkeit liegt unterhalb einer gewissen Schwelle. Je komplexer das zu lösende Problem jedoch ist, desto wahrscheinlicher ist es auch, daß unser Programm es nicht zufriedenstellend lösen wird: Wenn wir Glück haben, produziert es immerhin eine falsche Lösung; in der Regel führen komplexe Probleme aber zum ‘Hängenbleiben’ oder gar zum ‘Absturz’ des Programms. Das Problem, mit dem wir es hier zu tun bekommen, ist nun keineswegs mehr symbolischer Natur, vielmehr ist es in der Annahme begründet, wir könnten mit jeder uns zur Verfügung stehenden Hardware beliebig genaue Arithmetik mit reellen Zahlen betreiben. Nehmen wir an, unsere Hardware bestände aus Papier und Bleistift: Unter der Voraussetzung, daß uns ‘unendlich’ viel Zeit zur Verfügung stände, könnten wir mit Hilfe des besprochenen Verfahrens tatsächlich *jedes* allgemeine Flächenverschneidungsproblem lösen. Treffen wir dagegen die realistischere Annahme, uns stände nur ein elektronischer Rechner (und ein begrenzter Zeitraum) zur Verfügung, dann stehen wir vor dem Problem, daß notwendige Rechenoperationen eben nicht beliebig genau durchführbar sind: Die Tatsache, daß in elektronischen Rechnern jede Zahl, also auch jede reelle Zahl, in Form einer *endlichen* Folge elektronischer Zustände codiert wird, bedeutet für uns nicht mehr und nicht weniger, als daß jede Kommazahl nur eine *endliche* Anzahl von Kommastellen aufweist.

Man beachte, daß mit elektronischen Rechnern die arithmetischen Operationen Addition, Subtraktion und Multiplikation genau dann ‘unendlich genau’ durchführbar sind, wenn die verarbeiteten Zahlen alle ganzzahlig sind. Eine erste Strategie zur Vermeidung von Ungenauigkeiten bei unserem Verschneidealgorithmus könnte daher sein, die zu verarbeitenden Koordinaten — z.B. durch die Multiplikation mit einer großen Zahl wie  $10^{16}$  und anschließendem ‘Abschneiden’ der Kommastellen — so zu transformieren, daß sie alle ganzzahlig sind. Leider ist damit das Problem keineswegs aus der Welt geschafft, denn auch, wenn die Koordinaten der Endpunkte aller Kanten ganzzahlig sind, gilt dies nicht zwingend für die Schnittpunkte der Kanten (Abb. 3.21 links). Hinzu kommt, daß die Bestimmung des Schnittpunktes zweier Kanten die Durchführung einer Division erfordert (Kap. 3.5.1.1); das Ergebnis dieser Division kann wiederum nur mit endlicher Genauigkeit repräsentiert werden, ist aber dennoch die Grundlage für alle nachfolgenden Rechenoperationen. Mit anderen Worten: Auch wenn wir die Ausgangsdaten mit beliebig großer Genauigkeit darstellen könnten, werden wir irgendwann von Ungenauigkeiten in Gestalt von Schnittpunktkoordinaten eingeholt und müssen damit rechnen, daß die Fehler im weiteren Verlauf ‘verschleppt’ werden und sich ‘potenzieren’.

Zur Veranschaulichung betrachten wir Abb. 3.20. Dargestellt ist das bereits bekannte Beispiel aus Kap. 4.2, das das Streudiagramm der standardisierten Variablen ‘Tertiärisierungsgrad’ ( $x$ -Achse) und ‘Anteil der Frauen an den Arbeitslosen’ ( $y$ -Achse) für die 327 westdeutschen Kreise und kreisfreien Städte zeigt, wobei jede Raumeinheit nicht, wie üblich, als Punkt, sondern als Kreisfläche mit einem



**Abbildung 3.20** Verschiedene Vergrößerungen von 327 Zwölfеcken. Bei der Verschneidung beliebig vieler planarer Subdivisionen treten häufig Situationen auf, die eine Rechengenauigkeit erfordern, die mit der verwendeten Hardware nicht erzielt werden kann. Degenerierte Fälle, wie sich überlappende Kanten oder Knoten, in denen sich mehr als zwei Kanten schneiden, bilden die seltene Ausnahme; in der Regel handelt es sich um sich *fast* überlappende Kanten und *beinahe* aufeinanderliegende Knoten. Die Länge der kürzesten Dreiecksseite im Ausschnitt unten rechts beträgt ca. 0,0000069.

Radius von 0,25 dargestellt ist.<sup>23</sup> Jede Kreisfläche ist der Einfachheit halber als Zwölfeck approximiert. Wenn wir die durch diese Zwölfecke induzierte planare Subdivision bestimmen wollen, müssen wir im wesentlichen mit Hilfe des Scanline-Verfahrens die Menge aller Punkte bestimmen, in denen sich die Kanten aller Zwölfecke schneiden und an diesen Punkten in einer vorgegebenen Reihenfolge weitere symbolische Operationen durchführen (Kap. 3.2- 3.4). Insbesondere muß im Verlauf des ‘Scanline-Schwenks’ für jeden Eventpunkt die Statusfolge  $\mathcal{T}$  bekannt sein, in der die Reihenfolge der Kanten, die die aktuelle Scanlinie schneiden, von links nach rechts eindeutig festgelegt ist. Die Statusfolge ist als balancierter Binärbaum implementiert, so daß Kanten eingefügt, gelöscht und linke und rechte Nachbarn gesucht werden können. Notwendig ist dafür eine Vergleichsoperation, die für jedes Kantenpaar die (lokale) Ordnung dieses Paares angibt. Wir haben bereits gesehen, daß Überlappungsfälle auftreten können, so daß zwei Kanten in  $\mathcal{T}$  nicht unterscheidbar sind (Kap. 3.4). Was aber ist, wenn zwei Kanten, die sich in Wirklichkeit nicht überlappen, aus technischen Gründen nicht eindeutig unterscheidbar sind? Die Folge wäre der Versuch, eine (nicht wirklich vorliegende) Überlappung aufzulösen. In unserem Beispiel (Abb. 3.20) liegen zahlreiche solcher ‘Fast-Überlappungen’ vor, die sich erst bei 64-facher Vergrößerung ( $2^6$ ) zu erkennen geben. Selbst bei dieser Vergrößerung scheint die Differenz noch vernachlässigbar; erst bei 4096-facher Vergrößerung ( $2^{12}$ ) sieht man plötzlich, daß sich mit einer Vernachlässigung des Abstands der beiden sich anscheinend überlappenden Kanten die topologischen Eigenschaften der Knoten und Kanten erheblich ‘verbiegen’ würden! Vergrößert man schließlich 65536-fach ( $2^{16}$ ), muß man sogar feststellen, daß sich die drei Kanten  $e_1$ ,  $e_2$  und  $e_3$  keineswegs in einem gemeinsamen Punkt schneiden; die längste Seite des durch die drei Schnittpunkte gebildeten Dreiecks ist jedoch nur ganze  $6,7 \times 10^{-6}$  Standardabweichungen lang.

Können wir nun wirklich sicher sein, daß die Rechengenauigkeit groß genug ist, um

1. die Schnittpunkte, wenn auch nicht ‘unendlich’ genau, so doch wenigstens so genau zu berechnen, daß sie in der Eventfolge  $\mathcal{Q}$  in der symbolisch korrekten Reihenfolge  $e_2 \cup e_3 \prec_{\mathcal{Q}} e_1 \cup e_3 \prec_{\mathcal{Q}} e_1 \cup e_2$  eingefügt werden, und
2. bei der Bearbeitung des Eventpunktes  $e_1 \cup e_3$  die (z.T. extrem kurzen Reststücke der) Kanten in der Statusfolge  $\mathcal{T}$  in der symbolisch korrekten Reihenfolge  $e_2 \prec_{\mathcal{T}} e_1 \prec_{\mathcal{T}} e_3$  zu ordnen?

Reicht die Rechengenauigkeit dafür nicht aus, dann begeben wir uns gleichsam in eine Welt voller topologischer Widersprüche: Punkte, die in Wahrheit links von einer Kante liegen, liegen jetzt rechts von ihr, zwei Kanten überlappen sich nicht, aber schneiden sich in mehr als einem Punkt, eine Kante hat sich selbst als Nachbarn, der rechte Nachbar einer Kante ist ein linker Nachbar ihres linken Nachbarn — die Welt, die durch Rechengenauigkeiten entsteht, ist alles andere als logisch und führt beinahe zwangsläufig zu Situationen, in denen wir Kanten in der Statusfolge  $\mathcal{T}$  nicht wiederfinden, falsche Kanten aus  $\mathcal{T}$  löschen oder gar auf der Suche nach einer Kante in  $\mathcal{T}$  in eine ‘Endlosschleife’ geraten. Diese und ähnliche Situationen führen zu den besagten Ausgabefehlern, den Programmabbrüchen und dem ‘Hängenbleiben’.

Algorithmen, die auf die exakte Verarbeitung rationaler Zahlen angewiesen sind, aber so implementiert sind, daß sie trotz unzureichender Rechengenauigkeit in jedem Fall zu validen Ergebnissen führen, werden als *robust* bezeichnet (vgl. zusammenfassend Yap 1997). In diesem Sinne ist unser Algorithmus also nicht robust, und wir müssen untersuchen, für welche arithmetischen Operationen Rechenfehler von Bedeutung sind und welche Konsequenzen sich für unser Verfahren ergeben, wenn wir uns an diesen Stellen ‘verrechnen’. Genauer: Welche Modifikationen müssen wir durchführen, um den vorgestellten Algorithmus in einen robusten Algorithmus zu transformieren, der unabhängig von der möglichen Rechengenauigkeit eine valide Lösung für *jedes* allgemeine Flächenverschneidungsproblem angibt?

<sup>23</sup>Der Wert entspricht genau einem Viertel der Standardabweichung, da beide Variablen standardisiert sind.

### 3.5.1 Primitive arithmetische Operationen

Bisher wurde vorausgesetzt, für Probleme wie die Bestimmung eines Schnittpunktes oder die Bestimmung eines linken Nachbarn gäbe es jeweils eine primitive Lösungsoperation. Nun sind leider genau diese primitiven (*Rechen-*)Operationen die eigentliche Ursache für die Existenz von Widersprüchen im oben beschriebenen Sinne, da mit ihnen Informationen direkt anhand von bekannten (aber ggf. ungenauen) Knotenkoordinaten abgeleitet werden. Sie können folgendermaßen systematisiert werden:

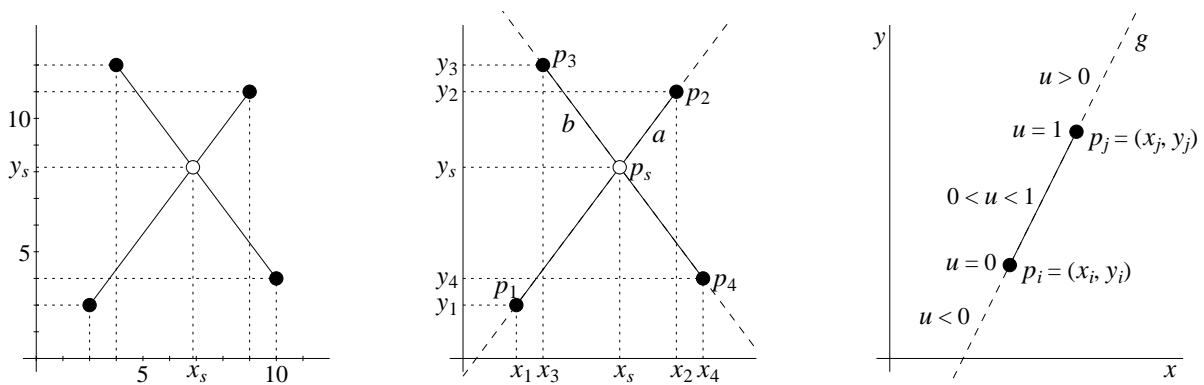
- Die Bestimmung des Schnittpunktes zweier Kanten beruht im wesentlichen auf der Gleichsetzung zweier Geradengleichungen, die mit den Endpunkten der Kanten aufgestellt werden können. Jede Schnittpunktbestimmung erfordert die Durchführung einer Division und ist gewissermaßen der Ursprung arithmetischer Ungenauigkeiten (Kap. 3.5.1.1).
- In der Eventfolge  $\mathcal{Q}$  muß die Ordnung  $\prec_{\mathcal{Q}}$  für jedes beliebige Paar von Eventpunkten eindeutig sein (vgl. S. 58); die Unterscheidung der Eventpunkte ist trivial, erfolgt aber aufgrund ihrer Koordinaten.
- In der Statusfolge  $\mathcal{T}$  muß die Reihenfolge zweier Kanten, die die aktuelle Scanlinie schneiden, im Sinne der Ordnung  $\prec_{\mathcal{T}}$  eindeutig sein (vgl. S. 58). Wie man leicht sieht, läßt sich jeder Vergleich entweder auf den Vergleich zweier Knoten reduzieren, die beide auf der Scanlinie liegen, oder auf die Bestimmung der relativen Lage der einen Kante zu dem Punkt, in dem die andere Kante die Scanlinie schneidet (Kap. 3.5.1.2). Beides basiert letztendlich auf den Koordinaten der Knoten bzw. Schnittpunkte.
- Bei der Aktualisierung der Flächeninformationen an einem beliebigen Eventpunkt muß die zyklische Ordnung aller Kanten, die an den Eventpunkt gebunden sind, eindeutig sein. Diese Ordnung kann anhand des Winkels bestimmt werden, den die Kanten mit der Horizontalen einschließen (vgl. S. 67). Ein ähnlicher Vergleich ist auch für zwei Segmente der Statusfolge erforderlich, wenn sie die Scanlinie im selben Punkt schneiden (ausschlaggebend ist dann ja immer die Reihenfolge unterhalb der Scanlinie). Die Bestimmung des Winkels würde jedoch die Berechnung einer trigonometrischen Funktion (z.B. der Cosinusfunktion) erfordern, was zeitaufwendig und erst recht anfällig für Rechenfehler ist. Auch die zyklischen Vergleiche lassen sich auf einfache Weise immer auf die Bestimmung der relativen Lage eines Knotens und einer Kante reduzieren (Kap. 3.5.1.2).

Im wesentlichen sind es also zwei Operationen, für die die Problematik der begrenzten Rechengenauigkeit von Bedeutung ist: einerseits die Bestimmung des Schnittpunktes zweier Kanten (hier werden die Rechenfehler ‘produziert’), andererseits die topologische Sortierung von Knoten und Kanten nach unterschiedlichen Ordnungen (hier entfalten Rechenfehler ihre fatale Wirkung).

#### 3.5.1.1 Berechnung des Schnittpunkts zweier Liniensegmente

Betrachten wir Abb. 3.21 (Mitte). Die beiden Liniensegmente (Kanten)  $a = (p_1, p_2)$  und  $b = (p_3, p_4)$  schneiden sich nur dann, wenn sich auch die Geraden schneiden, die durch  $p_1$  und  $p_2$  bzw. durch  $p_3$  und  $p_4$  verlaufen und wenn der Geradenschnittpunkt sowohl in  $a$  als auch in  $b$  liegt. Diese triviale Erkenntnis erlaubt es, den Schnittpunkt der Segmente  $a$  und  $b$  anhand der Parameterformen der Geradengleichungen zu bestimmen (Abb. 3.21 rechts; vgl. Laurini, Thompson 1995, S. 264ff.). Mit der Parametergleichung einer Geraden  $g$ , die durch die Punkte  $p_i$  und  $p_j$  verläuft, wird jeder Punkt  $p \in g$  beschrieben durch

$$p = p_i + u(p_j - p_i)$$



**Abbildung 3.21** Die Bestimmung des Schnittpunktes zweier Liniensegmente. Auch wenn die Koordinaten der Endpunkte zweier Kanten ganzzahlig sind, gilt dies nicht zwingend für ihren Schnittpunkt (links). Um den Schnittpunkt  $p$  der beiden Liniensegmente  $a$  und  $b$  zu bestimmen (Mitte), werden die Parameterformen der Geradengleichungen benutzt, die aus den Endpunkten  $p_i$  und  $p_j$  jedes Segments abgeleitet werden können (rechts).

bzw. in der Koordinatenschreibweise

$$\begin{aligned} x &= x_i + u(x_j - x_i) \\ y &= y_i + u(y_j - y_i) \end{aligned}$$

Der Parameter  $u$  gibt dabei für jeden beliebigen Punkt  $p \in g$  an, um welchen (relativen) Betrag er von  $p_i$  aus in Richtung  $p_j$  ‘verschoben’ ist. Im Fall von  $u = 0$  ist  $p$  gar nicht ‘verschoben’ ( $p = p_i$ ); im Fall von  $u = 1$  ist  $p$  ‘vollständig’ zu  $p_j$  ‘verschoben’ ( $p = p_j$ ). Für jeden Punkt  $p \in g$ , der zwischen  $p_i$  und  $p_j$  liegt, gilt daher  $0 < u < 1$ , d.h. für jeden Punkt, der auf dem Liniensegment  $\overline{p_i p_j}$  liegt, ist  $0 \leq u \leq 1$ .

Für die zwei Liniensegmente  $a$  und  $b$  in Abb. 3.21 (Mitte) sind die Parameterformen der entsprechenden Geradengleichungen also

$$\begin{aligned} p_a &= p_1 + u_a(p_2 - p_1) \\ p_b &= p_3 + u_b(p_4 - p_3) \end{aligned} \tag{3.1}$$

Löst man (3.1) für den Schnittpunkt  $p_s = p_a = p_b$ , erhält man

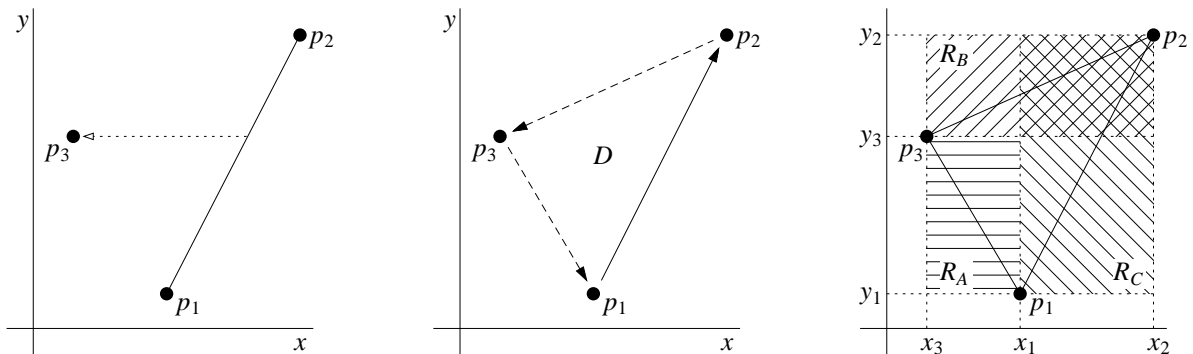
$$\begin{aligned} u_a &= \frac{(x_4 - x_3)(y_1 - y_3) - (x_1 - x_3)(y_4 - y_3)}{(x_2 - x_1)(y_4 - y_3) - (x_4 - x_3)(y_2 - y_1)} \\ u_b &= \frac{(x_2 - x_1)(y_1 - y_3) - (x_1 - x_3)(y_2 - y_1)}{(x_2 - x_1)(y_4 - y_3) - (x_4 - x_3)(y_2 - y_1)} \end{aligned} \tag{3.2}$$

Man beachte, daß die Nenner von  $u_a$  und  $u_b$  identisch sind. Insbesondere gilt:

- Wenn der Nenner 0 ist, sind die beiden Geraden parallel.
- Wenn sowohl die Zähler als auch der Nenner 0 sind, überlappen sich die beiden Geraden.

Für die beiden Liniensegmente  $a$  und  $b$  gilt mit Blick auf die Parameter  $u_a$  und  $u_b$  zusammenfassend:

- Wenn  $u_a = (0, 1)$  und  $u_b = (0, 1)$ , dann haben  $a$  und  $b$  einen gemeinsamen Endpunkt.
- Wenn  $u_a = (0, 1)$  und  $0 < u_b < 1$ , dann liegt ein Endpunkt von  $a$  auf  $b$ .



**Abbildung 3.22** Die Bestimmung der relativen Lage eines Punktes. Die Beobachtung, daß der Punkt  $p_3$  links vom Liniensegment  $\overline{p_1p_2}$  liegt, setzt implizit voraus, daß das Liniensegment gerichtet ist, d.h. ‘von unten nach oben’ verläuft (links). Unter dieser Voraussetzung läßt sich die relative Lage von  $p_3$  anhand der Orientierung des Dreiecks  $(p_1p_2p_3)$  ermitteln (Mitte). Dazu wird der Flächeninhalt des Dreiecks berechnet (rechts).

- Wenn  $0 < u_a < 1$  und  $u_b = (0, 1)$ , dann liegt ein Endpunkt von  $b$  auf  $a$ .
- Wenn  $0 < u_a < 1$  und  $0 < u_b < 1$ , dann schneiden sich  $a$  und  $b$ .

Durch Einsetzen eines der beiden Parameter in die entsprechende Geradengleichung aus (3.1) erhalten wir schließlich die Koordinaten des Schnittpunktes  $p_s = (x_s, y_s)$ , z.B. als

$$\begin{aligned} x_s &= x_1 + u_a (x_2 - x_1) \\ y_s &= y_1 + u_a (y_2 - y_1) \end{aligned} \tag{3.3}$$

### 3.5.1.2 Berechnung der relativen Lage eines Punktes

Die Lage eines Knotens relativ zu einer Kante (Abb. 3.22 links) ist immer dann von Bedeutung, wenn es darum geht, die Reihenfolge von Knoten bzw. Kanten in einer gegebenen Ordnung zu ermitteln. Die Ordnung selbst — sei es die Ordnung  $\prec_Q$  der Eventfolge  $Q$ , die Ordnung  $\prec_T$  der Statusfolge  $T$  oder die zyklische Ordnung aller Kanten um ihren gemeinsamen gerundeten Knoten — ist dabei unerheblich; alle besprochenen Ordnungen lassen sich leicht auf die Bestimmung der relativen Lage eines Knotens und einer Kante reduzieren. Das ist von großem Vorteil, da die relative Lage eines Punktes allein durch Summation und Multiplikation von Knotenkoordinaten ermittelt werden kann, so daß eine Entscheidung über die Reihenfolge zweier Kanten keine Division mit rationalen Zahlen erfordert und damit *relativ* stabil gegenüber Rechenfehlern ist.

Die Situation, die in Abb. 3.22 (links) dargestellt ist, wird jeder intuitiv so beschreiben, daß der Punkt  $p_3$  links vom Liniensegment  $\overline{p_1p_2}$  liegt. Das ist natürlich richtig, setzt aber gewissermaßen die Vogelperspektive voraus, nach der es ein ‘vorne’ (den oberen Blattrand) und ein ‘hinten’ (den unteren Blattrand) gibt: Wenn wir das Blatt umdrehen, liegt der Punkt  $p_3$  entsprechend rechts von  $\overline{p_1p_2}$ . Im kognitionspsychologischen Kontext mag diese Voraussetzung trivial erscheinen, im algorithmischen Kontext ist sie notwendige Voraussetzung für jede Differenzierung zwischen links und rechts (aber deshalb nicht weniger trivial). Mit anderen Worten: Mit der Beobachtung, daß  $p_3$  links von  $\overline{p_1p_2}$  liegt, wird implizit vorausgesetzt, daß das Liniensegment gerichtet ist, d.h. von  $p_1$  nach  $p_2$  verläuft und nicht umgekehrt (Abb. 3.22 Mitte). Die Frage, ob der Punkt  $p_3$  links oder rechts von  $\overline{p_1p_2}$  liegt, kann somit beantwortet werden, indem die Orientierung des Linienzuges  $(p_1p_2p_3p_1)$  ermittelt wird: Ist der

Linienzug gegen den Uhrzeigersinn orientiert, liegt  $p_3$  links, ist er dagegen im Uhrzeigersinn orientiert, liegt  $p_3$  rechts von  $\overline{p_1p_2}$ .<sup>24</sup>

Die Orientierung des Linienzuges, der das Dreieck  $D = (p_1p_2p_3)$  einschließt, kann nun sehr einfach bestimmt werden, indem der Flächeninhalt  $A_D$  von  $D$  (genau genommen das doppelte des Flächeninhalts) auf folgende Weise berechnet wird (Abb. 3.22 rechts; vgl. O'Rourke 1994): Zunächst wird das kleinste, das Dreieck  $D$  einschließende Rechteck  $R$  bestimmt, so daß der linke untere Eckpunkt von  $R$  der Punkt  $(\min(p_1.x, p_2.x, p_3.x), \min(p_1.y, p_2.y, p_3.y))$  und der rechte obere Eckpunkt von  $R$  entsprechend der Punkt  $(\max(p_1.x, p_2.x, p_3.x), \max(p_1.y, p_2.y, p_3.y))$  ist.  $D$  teilt  $R$  in maximal vier Flächen:  $D$  selbst und maximal drei weitere, äußere Dreiecke, die genau die Bereiche von  $R$  'ausfüllen', die nicht zu  $D$  gehören. Dem Flächeninhalt von  $D$  entspricht also der Flächeninhalt von  $R$  abzüglich der Summe aller Flächeninhalte dieser äußeren Dreiecke. Um die Größe der äußeren Dreiecke zu ermitteln, bestimmen wir nun die drei Rechtecke  $R_A$ ,  $R_B$  und  $R_C$ , so daß jede Dreiecksseite die Diagonale je eines dieser Rechtecke ist. Der Flächeninhalt eines äußeren Dreiecks entspricht also genau der Hälfte des Flächeninhalts von einem der drei Rechtecke. Sei  $A$  der Flächeninhalt von  $R$ , und seien  $A_A$ ,  $A_B$ ,  $A_C$  die Flächeninhalte von  $R_A$ ,  $R_B$ ,  $R_C$ . Der Flächeninhalt des Dreiecks  $D$  ist dann

$$A_D = A - \left( \frac{A_A}{2} + \frac{A_B}{2} + \frac{A_C}{2} \right)$$

mit

$$\begin{aligned} A &= (x_2 - x_3)(y_2 - y_1) \\ A_A &= (x_1 - x_3)(y_3 - y_1) \\ A_B &= (x_2 - x_3)(y_2 - y_3) \\ A_C &= (x_2 - x_1)(y_2 - y_1) \end{aligned}$$

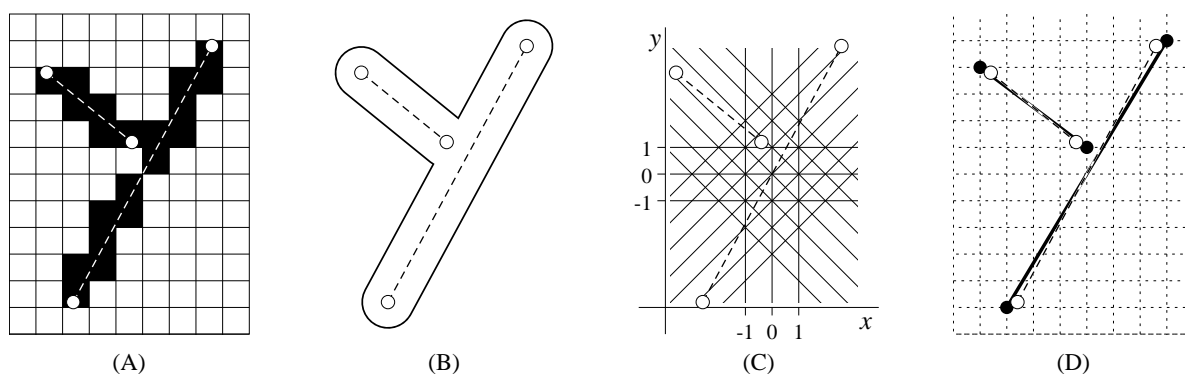
und der zweifache Flächeninhalt ist

$$\begin{aligned} 2A_D &= 2A - A_A - A_B - A_C \\ &= x_1 y_2 - x_1 y_3 + x_2 y_3 - x_2 y_1 + x_3 y_1 - x_3 y_2 \end{aligned} \quad (3.4)$$

Der Wert für  $A_D$  ist in unserem Beispiel positiv, denn der Linienzug, der das Dreieck  $D = (p_1p_2p_3)$  einschließt, ist gegen den Uhrzeigersinn orientiert und  $p_3$  liegt links von  $\overline{p_1p_2}$ . Dies muß nicht immer so sein, und wenn wir tatsächlich am Flächeninhalt von  $D$  interessiert wären, müßten wir streng genommen den Absolutwert von  $A_D$  bestimmen. Wie man leicht sieht, nimmt  $A_D$  nämlich negative Werte an, wenn  $p_3$  rechts von  $\overline{p_1p_2}$  liegt und der Linienzug im Uhrzeigersinn orientiert ist; verschieben wir  $p_3$  z.B. soweit nach rechts, daß  $x_3 > x_2$ , dann wären  $A$ ,  $A_A$  und  $A_B$  und damit auch  $A_D$  negativ. Zusammenfassend gilt daher:

$$A_D \begin{cases} < 0 & : p_3 \text{ liegt rechts von } \overline{p_1p_2} \\ 0 & : p_1, p_2, p_3 \text{ sind kollinear} \\ > 0 & : p_3 \text{ liegt links von } \overline{p_1p_2} \end{cases} \quad (3.5)$$

<sup>24</sup>Die Frage nach der Orientierung des genannten Linienzuges ist nicht zu verwechseln mit der Frage nach der Orientierung eines Halbkantenzuges in Kap. 3.2.3.2. Die Orientierung eines Halbkantenzuges wurde an dem am weitesten links liegenden Knoten  $v_l$  (dem Knoten mit der kleinsten  $x$ -Koordinate) festgestellt, indem der Winkel ermittelt wurde, der von den beiden durch  $v_l$  verlaufenden Halbkanten eingeschlossen wird. Genau genommen wird dabei natürlich kein Winkel berechnet; vielmehr ist durch die an die beiden Halbkanten gebundenen Knoten ein Dreieck gegeben, so daß es ausreicht, zu ermitteln, wie der Knoten  $v_l$  relativ zu der ihm gegenüberliegenden Dreiecksseite liegt.



**Abbildung 3.23** Die Konstruktion von Liniensegmenten fester Präzision. Die Ansätze zur Darstellung von Linien(-segmenten) fester Präzision reichen von einfacher *Diskretisierung* (A) über die *Intervall-Geometrie* (B) und die *gerundete Geometrie* (C) (dargestellt sind ideale Linien für  $K = 2$ ) bis hin zur *topologisch konsistenten Verzerrung* (D) (nach Yap 1997).

### 3.5.2 Repräsentationen mit fester Präzision

Wenn wir Vergleichsoperationen — z.B. bei der Bestimmung der Ordnung zweier Schnittpunkte (ist  $p_1.x < p_2.x$ ?) oder dem Test auf Kollinearität dreier Punkte (ist  $A_D = 0$ ?) — auf Grundlage numerischer Daten ausführen, die nur näherungsweise genau sind, dann existiert gewissermaßen ein Unsicherheitsbereich, innerhalb dessen exakte numerische Vergleiche keine validen Schlüsse zulassen. Für den robusten Umgang mit solchen ‘Rechenfehlern’ lassen sich zwei Strategien entwickeln (zusammenfassend Hoffmann 1989; Yap 1997): Zum einen kann man bei gegebener Präzision  $N$  des Inputs (quantifiziert als die Anzahl der zur Repräsentation einer numerischen Größe erforderlichen Bits) bestimmen, welche Präzision  $P = f(N)$  eine spezifische Rechenoperation erfordert, so daß die Präzision des Rechenergebnisses wiederum  $N$  ist. Wird jede Rechenoperation mit der Präzision  $P$  durchgeführt, sind Rechenfehler ausgeschlossen. Ein solcher Ansatz wird als ‘exakt’ bzw. ‘verzerrungsfrei’ bezeichnet. Für das Problem, den Schnittpunkt zweier Geraden anhand ihrer Parameterformen zu bestimmen, ist  $P = 3N + 3$  (Milenkovic 1989a); der Segmentschnitt erfordert  $P = 4N + 7$  (ebd.), und für die exakte Bestimmung der lokalen Ordnung  $\prec_{\mathcal{T}}$  ist in einigen Fällen  $P \geq 6N$  (ders. 1989b; vgl. auch Dobkin, Silver 1988). Angesichts der hohen Präzision, die erforderlich wäre, um das allgemeine Flächenverschneidungsproblem für  $m$  planare Subdivisionen  $\mathcal{S}_{1..m}$  zu lösen, ist dieser Ansatz hier eher von theoretischem Interesse. Zudem müssen wir uns fragen, ob ein exaktes Lösungsverfahren für unsere Zwecke überhaupt adäquat ist: Sind wir an derartig kleinen Flächen, wie sie z.B. Abb. 3.20 (unten rechts) zeigt, überhaupt interessiert? Die Antwort lautet nein, denn unser Interesse gilt nicht primär den geometrischen Details der angestrebten Lösung  $Sub(\mathcal{S}_{1..m})$ , sondern dem (visuellen) ‘Gesamteindruck’. Unser Ziel sollte daher vielmehr sein, nur hinreichend große Flächen in  $Sub(\mathcal{S}_{1..m})$  zu bilden; wünschenswert wäre zudem, daß wir die ‘minimale Größe’ dieser Flächen frei wählen können.

Wir wenden uns daher der zweiten Strategie zu, mit der versucht wird, die Komplexität des zu lösenden Problems und damit auch seiner Lösung zu verringern, indem sämtliche Rechen- und Vergleichsoperationen mit fester Präzision durchgeführt werden (Fortune 1993; Patrikalakis et al. 1995; Yap 1997). Die Daten (Koordinaten) werden also leicht verändert, d.h. die geometrischen Eigenschaften von  $\mathcal{S}_{1..m}$  und  $Sub(\mathcal{S}_{1..m})$  werden ‘verzerrt’. Im wesentlichen lassen sich vier Ansätze unterscheiden, und zwar am einfachsten anhand ihres unterschiedlichen Konzeptes einer Linie (vgl. Abb. 3.23):



- Bei der *Diskretisierung* besteht eine Linie — ähnlich wie in der klassischen Computergraphik — aus einer Menge benachbarter Pixel, wobei die Größe der Pixel umgekehrt proportional zur Präzision ist.<sup>25</sup> Wenn pixelartige Objekte im Fokus des Interesses stehen, mag ein solcher Ansatz adäquat sein. Für uns ist jedoch insbesondere wichtig, daß für die Flächen in  $Sub(\mathcal{S}_{1\dots m})$  über die Label der ‘Originalflächen’ der Rückgriff auf die ‘Originaldaten’ möglich ist. Dies kann mit einer ‘zellulären’ Struktur jedoch nicht gewährleistet werden.
- Die *Intervall-Geometrie* ist das geometrische Analogon zur Intervall-Arithmetik. Ein geometrisches Objekt wird dabei gewissermaßen zu einer Zone ‘verdickt’, die dieses Objekt enthält. Auf diese Weise wird — ähnlich wie bei einem mit GIS erzeugten Puffer — aus einem Punkt eine kreisförmige und aus einer Linie eine streifenförmige Fläche. Diese neuen Objekte werden als *Toleranzobjekte* (*toleranced objects*) bezeichnet; Toleranzobjekte, die sich berühren oder überschneiden, werden zu einem einzigen Objekt ‘verschmolzen’ (Segal, Sequin 1985).
- Mit dem Ansatz einer *gerundeten Geometrie* werden sog. *ideale Linien* gebildet. Ideale Linien werden durch die lineare Gleichung  $ax + by + c = 0$  beschrieben, wobei die Koeffizienten dieser Gleichung ganzzahlig und nach oben beschränkt sind:  $|a|, |b| < K, |c| < K^2$  für ein konstantes  $K$ . Linien, die dieser Gleichung entsprechen, werden als ‘repräsentierbar’ angesehen, und alle anderen Linien (bzw. Liniensegmente) werden auf diese repräsentierbaren Linien ‘gerundet’. Abb. 3.23 zeigt alle repräsentierbaren Linien für  $K = 2$  (Sugihara 1989; vgl. auch Hoffmann 1989).
- Analog zu idealen Linien werden bei einer *topologisch konsistenten Verzerrung* gewissermaßen ideale Punkte bestimmt, indem jeder Punkt (also auch jeder Endpunkt einer Linie und jeder Schnittpunkt zweier Linien) auf ein Gitter vorgegebener Maschenweite projiziert wird. Die Projektion eines Punktes auf ein diskretes Gitter erfolgt dadurch, daß die Punktkoordinaten auf die nächstliegenden Gitterkoordinaten ‘gerundet’ werden. Wie unten weiter ausgeführt wird, können durch das ‘Runden’ jedoch völlig neue Paare sich schneidender Linien entstehen, die sehr sorgfältig behandelt werden müssen, wenn topologische Konsistenz gewährleistet werden soll (Greene, Yao 1986; Hoffmann, Hopcroft, Karasick 1988; Hoffmann 1989; Milenkovic 1989a, 1989b; Fortune, Milenkovic 1991).<sup>26</sup>

Neben der schlichten Diskretisierung, die für unsere Zwecke leider gänzlich ungeeignet ist, ist der Ansatz einer topologisch konsistenten Verzerrung sicherlich am leichtesten nachvollziehbar. Zudem ermöglicht er weitestgehend die Beibehaltung aller bislang besprochenen Datenstrukturen und Verfahrensweisen: Wir haben es weiterhin mit Punkten, Linien und Flächen zu tun und können z.B. bei sehr nah aneinanderliegenden Kanten einen Lösungsweg beschreiten, der analog zum Auflösen sich (wirklich) überlappender Kanten ist. Im folgenden wird daher der Ansatz einer topologisch konsistenten Verzerrung verfolgt.

### 3.5.2.1 Modelle und Repräsentationen

Indem wir Koordinaten auf ein diskretes Gitter runden, wird jede (‘wahre und unendlich genaue’) rationale Zahl durch eine andere (‘falsche und nur endlich genaue’) Zahl repräsentiert, d.h. wir schränken die (infinite) Menge der wahren Koordinaten in der Euklidischen Ebene ein auf die (diskrete)

<sup>25</sup>Dieser Ansatz ist u.a. mit dem verbreiteten GIS ArcView implementiert.

<sup>26</sup>Strenggenommen werden von den genannten Autoren keine Punktkoordinaten gerundet, sondern analog zu den idealen Linien die Parameter der Gleichung  $ax + by + c = 0$ , und zwar so, daß die damit beschriebene Gerade durch zwei relevante Gitterpunkte verläuft. Da wir es beim allgemeinen Flächenverschneidungsproblem jedoch niemals mit Geraden zu tun haben, sondern immer mit Liniensegmenten (Strecken), ist die Vorstellung gerundeter Koordinaten völlig hinreichend.

Menge der repräsentierbaren aber unpräzisen Koordinaten, die mit einer Computerimplementation bearbeitet werden können. In diesem Sinne bearbeiten wir faktisch die *Repräsentation* eines fiktiven (da ‘unendlich genauen’) geometrischen *Modells* (Hoffmann, Hopcroft, Karasick 1988; Hoffmann 1989). Bezeichnen wir die ‘Maschenweite’ des diskreten Gitters mit  $d$ , so ist für zwei verschiedene repräsentierbare Zahlen  $z_1$  und  $z_2$   $|z_1 - z_2| \geq d$ . Ist eine Repräsentation ein Modell von sich selbst, sprechen wir von einem *natürlichen* Modell; z.B. ist das Dreieck  $((1,1)(3,2)(2,3))$  ein natürliches Modell für jedes  $d = (1, 0.5, 0.25, \dots)$ , da jede wahre Dreiecksordinate gleich ihrer repräsentierbaren Koordinate ist.

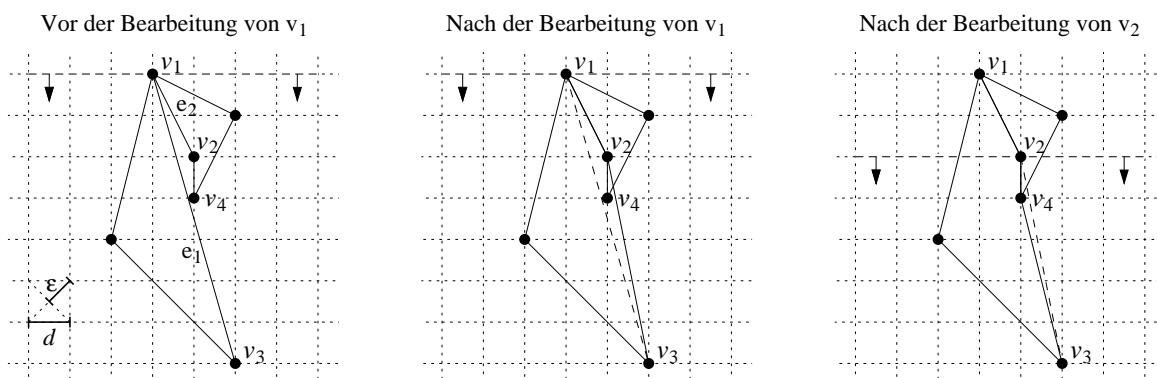
Eine binäre Operation  $op$ , z.B. die Verschneidung zweier Polygone, ist für zwei Repräsentationen  $R_1$  und  $R_2$  *korrekt* implementiert, wenn das Ergebnis von  $op(R_1, R_2)$  eine Repräsentation  $R_3$  ist und die Modelle  $M_1$ ,  $M_2$  und  $M_3$  existieren, so daß  $M_i$  das Modell von  $R_i$  ist und  $M_3 = op(M_1, M_2)$ . Wir sagen, ein Modell  $M$  einer gegebenen Repräsentation  $R$  ist zu  $R$   $\varepsilon$ -nah, wenn die symbolischen Informationen in  $M$  denen in  $R$  entsprechen und die größte numerische Abweichung der Knotenkoordinaten der Repräsentation von den exakten Knotenkoordinaten des Modells nicht größer als  $\varepsilon$  ist.<sup>27</sup> Wenn die Gitterweite des orthogonalen Rasters (die kleinstmögliche absolute Differenz zweier verschiedener repräsentierbarer Zahlen)  $d$  ist, dann entspricht  $\varepsilon$  der Hälfte der Diagonalen eines Gitterquadrats, d.h.  $\varepsilon = \frac{d}{\sqrt{2}}$  (vgl. Abb. 3.24 links). Jede  $m$ -näre Operation  $op$ , z.B. die Verschneidung von  $m$  Polygonen, ist dementsprechend genau dann korrekt implementiert, wenn für jede Repräsentation  $R_i$  ein Modell  $M_i$  existiert, der Algorithmus in jedem Fall eine Repräsentation  $R$  erzeugt und ein Modell  $M$  von  $R$  existiert, so daß  $M = op(M_1, M_2, \dots, M_m)$ . Die Präzision des Algorithmus kann dann wie folgt angegeben werden: Vorausgesetzt, jedes Modell  $M_i$  ist  $\varepsilon$ -nah zu seiner Repräsentation  $R_i$ , dann ist das Modell  $M$   $\delta(\varepsilon)$ -nah zu  $R$ . Unser Ziel ist daher, einen Algorithmus anzugeben, der das Modell  $M$  für alle legitimen Eingabedaten erzeugen kann, so daß die Funktion  $\delta$  existiert und  $\delta(\varepsilon)$  nicht übermäßig groß im Vergleich zu  $\varepsilon$  ist.

### 3.5.2.2 Symbolisches Schlußfolgern mit unpräziser Arithmetik

Die Implementation eines Algorithmus, der mit unpräzisen numerischen Daten operiert, führt letztendlich immer zum wiederholten Auftreten des folgenden Szenarios: Eine numerische Operation wird ausgeführt, und je nachdem, ob das Ergebnis  $r$  positiv, negativ oder Null ist, muß eine Entscheidung getroffen werden, die alle folgenden Operationen nachhaltig beeinflusst. Solange  $|r|$  größer als ein bestimmter Schwellenwert  $\tau$  ist, kann die Entscheidung mit Sicherheit getroffen werden. Ist dagegen  $|r| < \tau$ , dann ist eine Entscheidung, die sich auf  $r$  allein stützt, unsicher. Wenn die Entscheidung unsicher ist, könnten wir sie dem Zufall überlassen, oder wir treffen eine eindeutige Entscheidung, indem wir jedes Resultat  $|r| < \tau$  interpretieren als  $r = 0$ . Was immer wir an dieser Stelle tun, wir werden in jedem Fall unsere Geometrie ‘verbiegen’ und laufen Gefahr, uns in Widersprüche zu verstricken (Hoffmann 1989).

Abb. 3.24 zeigt ein Beispiel, das typisch für das Problem der Flächenverschneidung ist (ein weiteres einfaches Beispiel gibt Milenkovic 1989b). Die beiden dargestellten Flächen überschneiden sich nicht, nur die Begrenzungen berühren sich im Knoten  $v$  (links). Wenn  $v_1$  der aktuelle Eventpunkt ist, müssen alle Kanten, die an  $v_1$  gebunden sind und deren oberer Endpunkt in  $v_1$  liegt, in die Statusfolge  $\mathcal{T}$  eingefügt werden. Nehmen wir an, die Kante  $e_1$  wäre bereits in  $\mathcal{T}$  und die Kante  $e_2$  solle in  $\mathcal{T}$  eingefügt werden. In diesem Fall muß in  $v_1$  die lokale Ordnung  $\prec_{\mathcal{T}}$  für  $e_1$  und  $e_2$  bestimmt werden, was wiederum die Berechnung der Fläche  $A_D$  des Dreiecks  $D = (v_1, v_2, v_3)$  nach (Gleichung 3.4) erfordert. Wenn nun  $|A_D|$  unterhalb eines gewählten Schwellenwertes liegt, nehmen wir  $A_D = 0$  an, was für uns bedeutet, daß die Knoten  $v_1$ ,  $v_2$  und  $v_3$  kollinear sind und sich die Kanten  $e_1$  und  $e_2$  überlappen. Die

<sup>27</sup>Diese absolute Fehlernotation ist für unsere Zwecke ausreichend, obwohl sie natürlich bei Bedarf durch eine relative Notation ersetzt werden kann.



**Abbildung 3.24** Mögliche Folgen unsicherer numerischer Daten. Wenn  $v_1$  der aktuelle Eventpunkt ist, dann müssen die Kanten  $e_1$  und  $e_2$  (neben weiteren Kanten) in die Statusfolge  $\mathcal{T}$  eingefügt werden (links). Sind  $e_1$  und  $e_2$  aufgrund unpräziser Arithmetik in  $\mathcal{T}$  nicht unterscheidbar, werden sie als sich überlappend angesehen. Die Auflösung dieser Überlappung führt jedoch zur Überschneidung der beiden Flächen (Mitte), die in diesem Fall jedoch wieder ‘verschwindet’, wenn die Scanlinie den Knoten  $v_2$  erreicht (rechts).

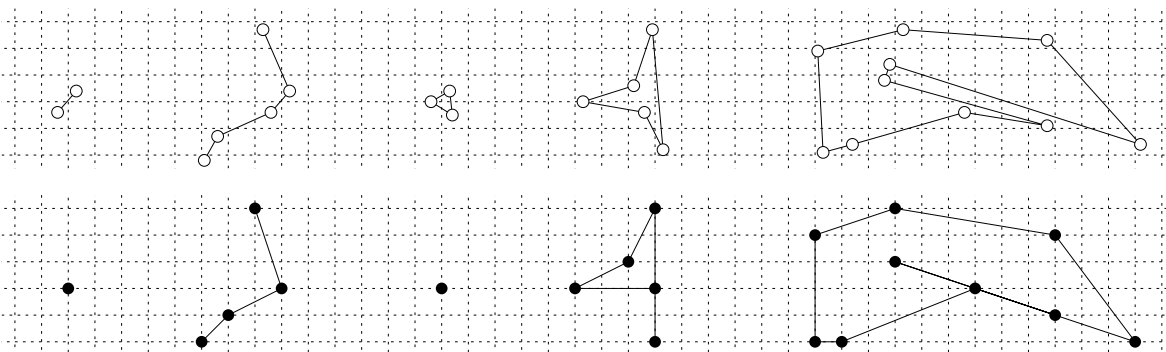
Überlappung wird aufgelöst, indem  $e_1$  im Knoten  $v_2$  geteilt wird. Das Resultat ist nun nicht bloß, daß die beiden Flächen eine gemeinsame Grenze aufweisen, sondern sie überschneiden sich sogar (Mitte)! Wie man leicht sieht, wird in diesem Fall die Flächenüberschneidung wieder aufgelöst, sobald die Scanlinie den Knoten  $v_2$  erreicht, denn es wird erneut eine (Fast-)Überlappung entdeckt (rechts); andere Situationen sind jedoch denkbar, in denen der Umgang mit derartigen ‘Rechenfehlern’ nicht so glatt verläuft wie hier.

Nehmen wir an, die linke Fläche in Abb. 3.24 sei die Repräsentation  $R_1$  des Modells  $M_1$ , die rechte Fläche sei die Repräsentation  $R_2$  des Modells  $M_2$ , und das Ergebnis der Verschneidung sei die Repräsentation  $R = (R_1 \cap R_2)$  des Modells  $M = (M_1 \cap M_2)$ .<sup>28</sup> In diesem Beispiel existiert während des Verfahrens ein Widerspruch zwischen der Repräsentation  $R$  und ihrem Modell  $M$ : Während sich in  $M$  keine Fläche überschneidet, existiert in  $R$  eine Schnittfläche. Anders ausgedrückt: Der Knoten  $v_4$  liegt in  $M$  links, in  $R$  dagegen zwischenzeitlich rechts von  $e_1$ , so daß ein Schnittpunkt in  $R$  existiert, der in  $M$  unmöglich ist. Die Tatsache, daß dieser Widerspruch nicht zum Abbruch des Verfahrens führt, verdanken wir einzig und allein dem Umstand, daß der ‘falsche Schnittpunkt’ unterhalb der Scanlinie liegt und ein Eventpunkt zwischen der aktuellen Scanlinie und diesem Schnittpunkt existiert, an dem er durch die erneute Auflösung einer (Fast-)Überlappung eliminiert wird. Läge der Schnittpunkt dagegen oberhalb der Scanlinie, würden wir ihn aller Wahrscheinlichkeit nach unberücksichtigt lassen, so daß als Ergebnis der Verschneidung (wenn wir überhaupt eines erhalten) eine ungültige Subdivision stände. Das Verfahren terminiert also in diesem Fall rein zufällig.

Bevor wir die Problematik genauer untersuchen, halten wir folgendes fest: Beim Umgang mit unpräzisen Daten können Widersprüche entstehen. Diese Widersprüche sind rein symbolischer Natur, und da Schlußfolgerungen aufgrund symbolischer Information für unseren Algorithmus von herausragender Bedeutung sind, haben Widersprüche für das weitere Verfahren ggf. schwerwiegende Konsequenzen. So muß z.B. für jeden Knoten die links benachbarte Kante bekannt sein; für den Knoten  $v_2$  in Abb. 3.24 (links) ist dies die Kante  $e_1$  (vgl. Kap. 3.2.3.3 und 3.3.2). Nachdem die erste Überlappung bearbeitet wurde, gilt diese Beziehung jedoch nicht mehr!

Je mehr wir auf symbolische Schlußfolgerungen angewiesen sind, desto schwieriger erweist sich die Aufgabe, einen robusten Verschneidealgorithmus anzugeben. Hoffmann, Hopcroft, Karasick (1988;

<sup>28</sup>Obwohl das dargestellte Verfahren noch nicht abgeschlossen ist, entspricht  $R$  der Darstellung in Abb. 3.24 (rechts), da unterhalb der Scanlinie keine weiteren Überlappungen oder Schnittpunkte existieren.



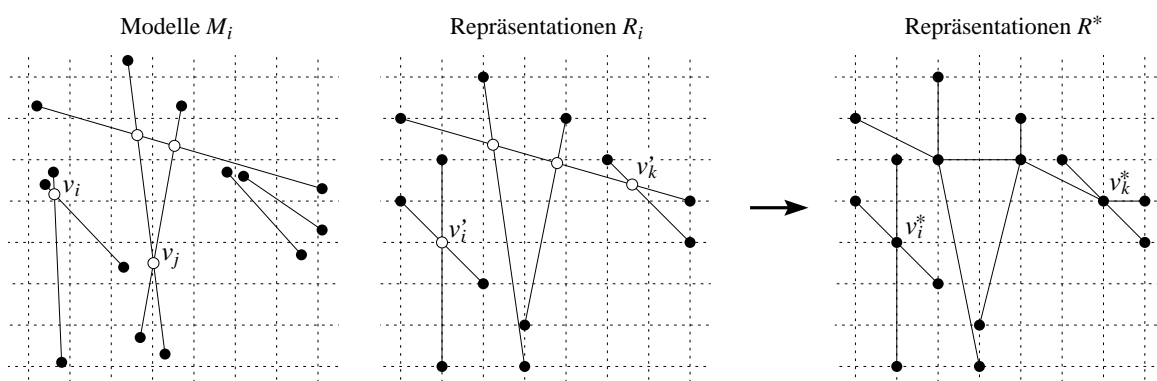
**Abbildung 3.25** Modelle und Repräsentationen. Verschiedene geometrische Modelle  $M$ , die ‘unendlich genaue’ Knotenkoordinaten aufweisen (obere Reihe), können in Form ihrer  $\varepsilon$ -nahen Repräsentationen  $R$  gewissermaßen ‘kollabieren’ (untere Reihe). Kanten in  $M$  können als Punkte in  $R$  repräsentiert sein, Flächen in  $M$  haben in  $R$  keine Ausdehnung, an Flächengrenzen entstehen ‘Sackgassen’ usf.

auch Hoffmann 1989) weisen z.B. darauf hin, daß bereits die simultane Verschneidung von drei Polygonen eine nur schwer lösbare Aufgabe darstellt, wenn der Algorithmus auf die Gültigkeit von Theoremen der darstellenden Geometrie sowohl für ein Modell  $M$  als auch für seine Repräsentation  $R$  angewiesen ist: Ein Schnittpunkt, der aufgrund des Pascalschen Theorems in  $M$  existieren muß, existiert in  $R$  z.B. nicht, wenn durch die Rundung der Knotenkoordinaten die Kollinearität der in Frage stehenden Knoten aufgehoben wurde.<sup>29</sup> Wenn verlangt wird, daß Knoten, die in  $M$  kollinear sind, auch in  $R$  kollinear sind, gestaltet sich jeder robuste Verschneidealgorithmus als derart komplex, daß die simultane Verschneidung von mehr als vier Polygonen noch nicht einmal angedacht wurde (ebd.; vgl. auch Abb. 3.26).

Mit dem Ergebnis der Verschneidung in Abb. 3.24 können wir allerdings zunächst zufrieden sein: In  $R$  existieren wie in  $M$  genau zwei Flächen, und die Tatsache, daß die Begrenzungen der beiden Flächen in  $R$  nun zwei Kanten und nicht nur, wie in  $M$ , einen Knoten gemeinsam haben, steht zu unserem Ziel nicht unbedingt im Widerspruch: Das Modell  $M$  soll bzw. kann ja seiner Repräsentation  $R$  nicht exakt entsprechen, sondern ist zu  $R$  eben nur  $\varepsilon$ -nah. Die in Kap. 3.5.2.1 geforderte Entsprechung der symbolischen Informationen in  $M$  und  $R$  ist daher auch zu relativieren: Ein Knoten  $v$ , der in  $M$  z.B. links von einer Kante  $e$  liegt, mag in  $R$  ebenfalls links von  $e$  oder auch auf  $e$  liegen, aber niemals rechts von  $e$ . Dabei nehmen wir nun nicht nur in Kauf, daß das Modell  $M$  von  $Sub(\mathcal{S}_{1\dots m})$   $\delta(\varepsilon)$ -nah zur Repräsentation  $R$  von  $Sub(\mathcal{S}_{1\dots m})$  ist, sondern auch, daß Knoten, Kanten und Flächen in  $M$  existieren können, die keine Entsprechung in  $R$  haben. Abb. 3.25 verdeutlicht diesen Sachverhalt: Verschiedene geometrische Konfigurationen, die ‘unendlich genaue’ Knotenkoordinaten aufweisen (obere Reihe) können durch die Projektion der Knotenkoordinaten auf die nächstgelegenen Gitterpunkte gewissermaßen ‘kollabieren’ (untere Reihe). Auf diese Weise können Kanten in  $M$  als Punkte (Kanten der Länge 0) in  $R$  repräsentiert sein, Flächen in  $M$  haben in  $R$  keine Ausdehnung, an Flächengrenzen entstehen ‘Sackgassen’ usf. Allgemein gilt also:

- Wenn für zwei Kanten  $e_1, e_2$  aus  $M$  die lokale Ordnungsrelation  $e_1 \prec_{\mathcal{T}} e_2$  gilt, dann gilt für ihre Repräsentationen  $e'_1, e'_2$  aus  $R$  — falls sie existieren — immer  $e'_1 \preceq_{\mathcal{T}} e'_2$ .
- Wenn für zwei Knoten  $v_1, v_2$  aus  $M$  die Ordnungsrelation  $v_1 \prec_{\mathcal{Q}} v_2$  in der Eventfolge  $\mathcal{Q}$  ist,

<sup>29</sup>Pascals Theorem: Wenn je drei Punkte  $v_1, v_3, v_5$  und  $v_2, v_4, v_6$  kollinear sind, dann sind die drei Schnittpunkte der Linien  $(v_1, v_2)$  und  $(v_4, v_5)$ ,  $(v_2, v_3)$  und  $(v_5, v_6)$  sowie  $(v_3, v_4)$  und  $(v_6, v_1)$  ebenfalls kollinear.



**Abbildung 3.26** Modelle und Repräsentationen von Schnittpunkten. Der Schnittpunkt  $v_j$  existiert als Modell (links), jedoch nicht als Repräsentation (Mitte). Der Schnittpunkt  $v'_k$  existiert dagegen als Repräsentation und hat keine Entsprechung als Modell. Die Abweichung des Schnittpunkts  $v_i$  von seiner Repräsentation  $v'_i$  nach der Verschneidung ist größer als  $\varepsilon$  (rechts).

dann ist die Ordnungsrelation ihrer Repräsentationen  $v'_1, v'_2$  aus  $R$  — falls sie existieren — immer  $v'_1 \preceq_Q v'_2$ .

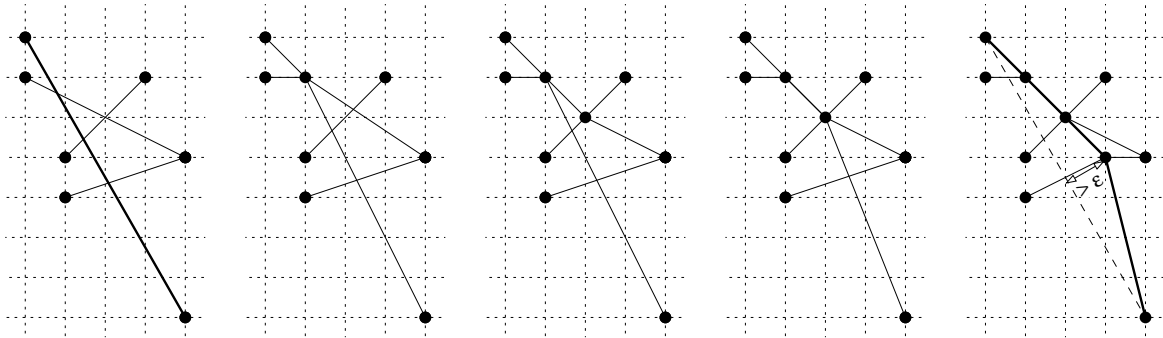
- Wenn sich zwei Flächen  $f_1, f_2$  aus  $M$  nicht überschneiden, dann sind auch ihre Repräsentationen  $f'_1, f'_2$  aus  $R$  — falls sie existieren — überschneidungsfrei.

### 3.5.2.3 Schnittpunktrepräsentationen und Präzision

Betrachten wir nun die Koordinaten von Schnittpunkten: Schnittpunkte liegen in der Regel nicht ‘automatisch’ auf den Gitterpunkten, sondern irgendwo dazwischen, und müssen daher auf das Raster der repräsentierbaren Koordinaten gerundet werden. Abb. 3.26 zeigt, was geschieht: Die Modelle  $M_i$  von sieben Liniensegmenten schneiden sich in vier verschiedenen Punkten (links). Die Projektion der Knotenkoordinaten aus den  $M_i$  auf die Menge der repräsentierbaren Koordinaten (die Gitterpunkte) führt zu sieben Repräsentationen  $R_i$  der Liniensegmente, die sich ebenfalls in vier Punkten schneiden (Mitte)<sup>30</sup>. Jedes Modell  $M_i$  ist  $\varepsilon$ -nah zu seiner Repräsentation  $R_i$ , d.h. die Distanz jedes Knotens aus  $M_i$  zu dessen Repräsentation in  $R_i$  ist nicht größer als  $\varepsilon$ . Dies gilt allerdings nicht für die Schnittpunkte: Die Distanz des Schnittpunkts  $v_i$  zweier Modelle zum Schnittpunkt  $v'_i$  der beiden entsprechenden Repräsentationen ist z.B. deutlich größer als  $\varepsilon$ . Und mehr noch: Es existieren zwei Modelle, die sich im Punkt  $v_j$  schneiden, deren Repräsentationen jedoch überschneidungsfrei sind; umgekehrt existieren zwei Repräsentationen, die sich im Punkt  $v'_k$  schneiden, deren Modelle jedoch keinen Punkt gemeinsam haben. Der Begriff der  $\varepsilon$ -Nähe kann sich demnach ausschließlich auf die in  $M_i$  und  $R_i$  bekannten Endpunkte von Liniensegmenten (Knoten) beziehen. Diese Einschränkung ist im Grunde genommen trivial, denn würden wir die Schnittpunkte sowohl der  $M_i$  als auch der  $R_i$  als Knoten interpretieren, setzten wir sie implizit als bekannt voraus — wären sie das, könnten wir uns die gesamte Diskussion ersparen.

Die Schnittpunkte der  $M_i$  sind daher gegenstandslos, und wir sind einzig an den Schnittpunkten der  $R_i$  interessiert. Jeder Schnittpunkt zweier Repräsentationen hat nun in der Regel Koordinaten, die wiederum nicht zur Menge der repräsentierbaren Koordinaten gehören (Abb. 3.26 Mitte). Die Koordinaten dieser Schnittpunkte müssen daher ebenfalls auf die Gitterpunkte gerundet werden, so daß jeder

<sup>30</sup>Im Punkt  $v'_k$  schneiden sich drei Liniensegmente, von denen sich zwei vollständig überlappen. Gehen wir davon aus, daß diese Überlappung aufgelöst wurde, liegen nur sechs Liniensegmente vor. Zwei Modelle entsprechen dann derselben Repräsentation.



**Abbildung 3.27** Die Präzision einer Verschneidung mit verzerrter Topologie. Wenn Schnittpunkte während des Scanline-Verfahrens auf ein diskretes Raster projiziert werden, sind die Kanten vor und nach der Teilung in der Regel nicht deckungsgleich. Wird eine Kante in  $R$  von mehreren anderen Kanten in  $R$  geschnitten, kann sich diese Ungleichheit gewissermaßen kumulativ verstärken.

‘wahre’ Schnittpunkt zweier Repräsentationen wiederum ein Modell darstellt, das im Ergebnis der Verschneidung  $\varepsilon$ -nah repräsentiert wird (rechts). Dieses Ergebnis besteht also aus Repräsentationen der  $R_i$  und ist damit gewissermaßen eine ‘Meta-Repräsentation’, die wir mit  $R^*$  bezeichnen.

Nun ist die Repräsentation  $R^*$  nicht, wie es nach Abb. 3.26 den Anschein hat, in jedem Fall  $\varepsilon$ -nah zu ihrem ‘Modell’  $R = (R_1, R_2, \dots, R_m)$ .  $R^*$  ist das Ergebnis einer mehrfach angewendeten Verschneideoperation, mit der jeder Schnittpunkt einzeln auf das Raster der repräsentierbaren Koordinaten projiziert wurde. Genauer:  $R^*$  ist die Repräsentation des planaren Graphen, der durch die Kanten in  $R$  induziert wird, d.h. jeder Knoten in  $R^*$  liegt auf einem Gitterpunkt. Bei jeder Verschneideoperation werden maximal zwei Kanten in einem gemeinsamen Punkt geteilt, so daß aus einer Kante maximal zwei neue Kanten entstehen. Diese neuen Kanten haben wir bislang als natürliche Modelle aufgefaßt, die deckungsgleich mit den alten Kanten sind (vgl. Kap. 3.2.3.1). Wenn die Schnittpunkte dagegen auf ein diskretes Raster projiziert werden, sind die Kanten vor und nach der Teilung in der Regel nicht deckungsgleich. Wird nun eine Kante in  $R$  von mehreren anderen Kanten in  $R$  geschnitten, kann sich diese Ungleichheit gewissermaßen kumulativ verstärken. Abb. 3.27 zeigt einen solchen Effekt: Die dicke gezeichnete Kante (ganz links) wird im Laufe eines Scanline-Schwenks immer weiter ‘verbogen’ und nach Abschluß des Verfahrens durch einen Kantenzug repräsentiert, der von der ursprünglichen Kante z.T. deutlich weiter als  $\varepsilon$  entfernt ist. Wir messen nun nicht mehr nur die Entfernung zwischen zwei Knoten, sondern die Entfernung zwischen einer Kante  $e$  aus  $R$  und einem Kantenzug  $E$  aus  $R^*$ , und zwar als maximale Länge aller Lotrechten, die von den Knoten in  $E$  auf  $e$  gefällt werden können. Diese Distanz entspricht in etwa der Funktion  $\delta(\varepsilon)$  (Kap. 3.5.2.1), für die wir jetzt eine Obergrenze angeben können:

**Hilfssatz 3.7.** Sei  $D$  eine Menge von rationalen Zahlen, für jede Zahl  $p \in Z$  und der nächstgrößeren Zahl  $s \in Z$  sei  $s - p = d$ . Sei  $\varepsilon = \frac{d}{\sqrt{2}}$ , sei  $R = \{e'_1, e'_2, \dots, e'_m\}$  eine Repräsentation von  $m$  Kanten, deren Knotenkoordinaten alle aus  $D$  sind, und sei  $M$  das  $\varepsilon$ -nahe Modell von  $R$ . Sei  $R^*$  die Repräsentation des planaren Graphen, der durch die Kanten in  $R$  induziert wird, so daß die Koordinaten jedes Knotens in  $R^*$  aus  $D$  sind. Jede Kante  $e_i$  aus  $M$  ist dann zu dem Kantenzug, durch den  $e_i$  in  $R^*$  repräsentiert wird,  $\delta(\varepsilon)$ -nah, so daß  $\delta(\varepsilon) \leq (k + 1)\varepsilon$ .

*Beweis.* Sei  $v'$  der Schnittpunkt zweier Kanten  $e'_i, e'_j$  aus  $R$  nach (Gleichung 3.3). Wenn  $e'_i$  in  $v'$  geteilt wird, entstehen maximal zwei neue Kanten  $e'_{i1}, e'_{i2}$ . Nehmen wir zunächst an,  $v'$  sei in  $R$  repräsentierbar, d.h.  $v.x, v.y \in D$ . In diesem Fall liegen die neuen Kanten exakt auf  $e'_i$ , und die Distanz zwischen

$e'_{i1}, e'_{i2}$  und  $e'_i$  ist 0. Wenn die Schnittpunkte aller Kanten aus  $R$  in  $R$  repräsentierbar sind, ist  $\delta(\varepsilon)$  gleich der maximalen Distanz zwischen einem Knoten aus  $M$  und einem Knoten aus  $R$ . Da  $M$   $\varepsilon$ -nah zu  $R$  ist, ist also  $\delta(\varepsilon) \leq \varepsilon$ .

Nehmen wir nun an, der Schnittpunkt  $v'$  sei in  $R$  nicht repräsentierbar, d.h.  $v'.x \notin D \vee v'.y \notin D$ . Die maximale absolute Differenz einer Koordinate von  $v'$  zur nächsten repräsentierbaren Zahl aus  $D$  ist  $\frac{d}{2}$ , also ist die maximale Distanz zwischen  $v'$  und dem nächsten in  $R$  repräsentierbaren Knoten  $v^*$   $\varepsilon = \frac{d}{\sqrt{2}}$ . Sei  $e'_i = (v'_u, v'_l)$  eine Kante aus  $R$ , die in  $v^*$  geteilt wird, wobei  $v'_u$  der obere Endpunkt von  $e'_i$  ist. Ist  $v^* = v'_u$  oder  $v^* = v'_l$ , bleibt  $e'_i$  ungeteilt. Ist  $v^* \neq v'_u \neq v'_l$ , entstehen durch die Teilung in  $v^*$  genau zwei neue Kanten  $e'_{i1} = (v'_u, v^*)$  und  $e'_{i2} = (v^*, v'_l)$ , die je einen Knoten mit  $e'_i$  gemeinsam haben. Die einfache Teilung der Kante  $e'_i$  führt daher zu einem Kantenzug von  $v'_u$  über  $v^*$  zu  $v'_l$ , dessen Distanz zu  $e'_i$  der Länge der Lotrechten entspricht, die von  $v^*$  auf  $e'_i$  gefällt werden kann. Da die Distanz zwischen  $v'$ , der ja auf  $e'_i$  liegt, und  $v^*$  nicht größer als  $\varepsilon$  ist, ist  $e'_i$  zu dem entstandenen Halbkantenzug  $\varepsilon$ -nah.

Da eine Kante nach ihrer Teilung nicht mehr in  $R$  existiert, wird jede Kante aus  $R$  höchstens einmal geteilt. Wird eine Kante  $e'_i = (v'_u, v'_l)$  aus  $R$  von mehr als einer anderen Kante aus  $R$  geschnitten, so wird  $e'_i$  also nur im ersten gefundenen Schnittpunkt  $v^*_1$  geteilt. Ein weiterer Schnittpunkt  $v^*_2$  führt entweder zur Teilung der Kante  $e'_{i1} = (v'_u, v^*_1)$  oder zur Teilung der Kante  $e'_{i2} = (v^*_1, v'_l)$ . Die geteilte Kante würde wiederum in  $R$  nicht weiterexistieren usw. Jeder gefundene Schnittpunkt führt also zur Teilung von maximal zwei Kanten aus  $R$ , jede Teilung einer Kante wiederum zu zwei neuen Kanten, die zur geteilten Kante  $\varepsilon$ -nah sind. Jede Teilung einer Kante führt daher zu einem Zuwachs der Distanz zwischen  $e'_i$  und ihrer momentanen Repräsentation in  $R$ , der durch  $\varepsilon$  nach oben beschränkt ist. Der Halbkantenzug, durch den  $e'_i$  der nach  $k$  Teilungen beim Abschluß des Verfahrens in  $R^*$  repräsentiert wird, kann demnach höchstens  $k\varepsilon$  sein.

Da  $M$  das  $\varepsilon$ -nahe Modell von  $R$  ist, ist die maximale Distanz zwischen jeder beliebigen Kante  $e_i$  aus  $M$  und ihre Repräsentation  $e'_i$  aus  $R$  durch  $\varepsilon$  nach oben beschränkt. Daher ist die maximale Distanz zwischen einer beliebigen Kante  $e_i$  aus  $M$  und dem Kantenzug, durch den  $e_i$  in  $R^*$  repräsentiert wird,  $k\varepsilon + \varepsilon$ , so daß  $\delta(\varepsilon) \leq (k + 1)\varepsilon$ .

□

### 3.5.3 Verschneidung mit topologisch konsistenter Verzerrung

Bei der Verschneidung der  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  haben wir über die Wahl der Gitterweite  $d$  die Möglichkeit, die Größe von  $\varepsilon$  und damit die Komplexität des Ergebnisses  $Sub(\mathcal{S}_{1\dots m})$  nach Belieben zu steuern: Je größer wir  $d$  wählen, desto weniger komplex ist das Ergebnis der Verschneidung. Wir wollen das Problem so strukturieren, daß wir zunächst nur die Knoten und Kanten des  $Sub(\mathcal{S}_{1\dots m})$  zugrundeliegenden planaren Graphen berechnen. Diese Berechnung kann als 'kontrolliertes Kollabieren' des Modells des Graphen von  $\bigcup_{i=1}^m \mathcal{S}_i$  betrachtet werden, und je größer wir das Gitter wählen, desto eher werden wir mit unliebsamen Folgen dieses Kollabierens konfrontiert. Auf zwei Dinge ist insbesondere zu achten: Wir müssen (1.) die Folgen einer verzerrten Topologie so kontrollieren, daß wir tatsächlich *alle* Schnittpunkte finden, also auch jene, die sich erst im Laufe des Verfahrens aufgrund der  $\varepsilon$ -nahen Repräsentation von Schnittpunkten ergeben, und (2.) die notwendigen symbolischen Informationen (im wesentlichen die Flächeninformationen) sammeln, um diese *nach* der Berechnung der Knoten und Kanten zur vollständigen doppelt-verketteten Kantenliste von  $Sub(\mathcal{S}_{1\dots m})$  'zusammenzusetzen'.

Bevor wir einen Knoten in die doppelt-verkettete Kantenliste  $\mathcal{D}$  und in die Eventfolge  $\mathcal{Q}$  einfügen, werden dessen Koordinaten zunächst auf das Raster der repräsentierbaren Koordinaten projiziert. Damit erzeugen wir bei der Initialisierung von  $\mathcal{D}$  für jedes Modell  $M_i$  einer planaren Subdivision  $\mathcal{S}_i$

dessen Repräsentationen  $R_i$ . Jedes Modell  $M_i$  ist  $\varepsilon$ -nah zu seiner Repräsentation  $R_i$ , und die symbolischen Informationen in  $M_i$  und  $R_i$  stimmen im obigen Sinne relativ überein. Bereits zu Beginn des eigentlichen Scanline-Schwenks hat also jeder Knoten in  $\mathcal{D}$  gerundete Koordinaten; entsprechend ist auch jeder Eventpunkt in der Eventfolge  $\mathcal{Q}$  repräsentierbar, so daß bei hinreichend großem  $\varepsilon$  alle Eventpunkte eindeutig unterscheidbar sind. Wenn wir  $\varepsilon$  so groß wählen, daß die Berechnung der kleinstmöglichen Dreiecksfläche nach (Gleichung 3.4), die von drei verschiedenen Eventpunkten aufgespannt werden kann, sicher ist, können wir sowohl jede lokale Ordnungsrelation in der Statusfolge  $\mathcal{T}$  als auch jede zirkuläre Ordnung aller an einen Knoten gebundenen Kanten sicher bestimmen.<sup>31</sup> Wir wenden dann das Scanlinie-Verfahren auf die Repräsentationen  $R_i$  an, so wie sie in der (zu Beginn des Verfahrens ‘ungültigen’) doppelt-verketteten Kantenliste  $\mathcal{D}$  vorliegen. Dabei ändern wir die Vorgehensweise, wie sie in Kap. 3.2 vorgestellt wurde, wie folgt:

- Mit jedem Eventpunkt  $v$  werden nicht nur diejenigen Kanten gespeichert, deren oberer Endpunkt in  $v$  liegt, sondern *alle* Kanten, die an  $v$  gebunden sind, und zwar in ihrer zyklischen Ordnung um  $v$ .
- Wenn ein Schnittpunkt  $v$  gefunden wird, werden die sich schneidenden Kanten, ähnlich wie bei der Auflösung von Überlappungen, *sofort* im nächsten repräsentierbaren Knoten  $v'$  geteilt.

### 3.5.3.1 Lokale Verletzungen der Invariante

So einfach, wie es die Abbildungen 3.26 und 3.27 erscheinen lassen, ist das Problem einer topologisch konsistent verzerrten Verschneidung leider nicht gelagert. Da wir dem Scanline-Prinzip folgen, ist die Reihenfolge, in der wir die Schnittpunkte finden, nicht von vorneherein bekannt; alles, was wir wissen, ist, daß neue Schnittpunkte ausschließlich unterhalb der Scanlinie gesucht werden müssen, und daß oberhalb der Scanlinie ‘alles in Ordnung’ ist. Diese Beobachtung hatten wir auf S. 61 als *Invariante des Scanline-Prinzips* bezeichnet und uns von da an immer wieder zunutze gemacht. Aufgrund der zentralen Bedeutung der Invariante sind wir gut beraten, wenn wir peinlich genau darauf achten, daß wir sie an keiner Stelle des Verfahrens verletzen. Wie sich zeigt, passiert jedoch genau das, wenn wir versuchen, das allgemeine Flächenverschneidungsproblem mit fester Präzision zu lösen.

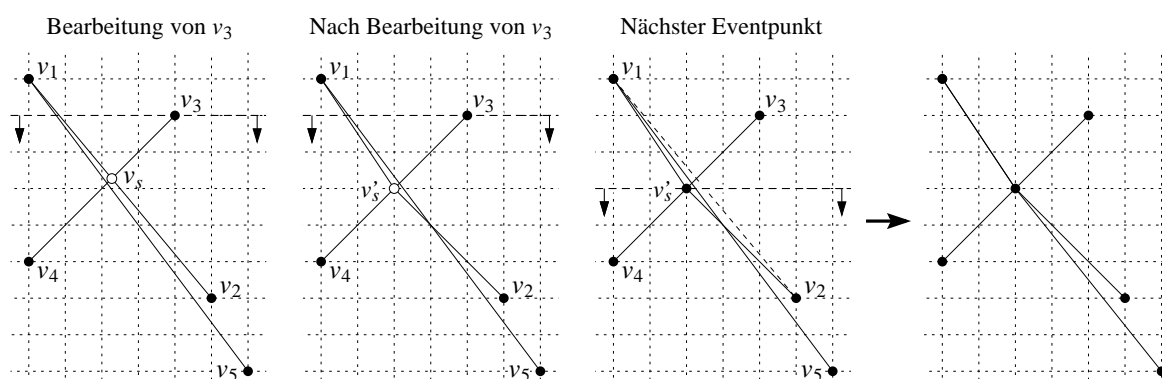
### Überziehungen

Betrachten wir Abb. 3.28.<sup>32</sup> Im aktuellen Eventpunkt  $v_3$  ist die lokale Ordnung der Kanten in der Statusfolge  $\mathcal{T}$   $(v_1, v_5) \prec_{\mathcal{T}} (v_1, v_2) \prec_{\mathcal{T}} (v_3, v_4)$  (links). Da  $(v_1, v_2)$  der linke Nachbar von  $(v_3, v_4)$  ist, werden diese beiden Kanten auf Intersektion geprüft (Zeile 13 in `BERABEITEEVENTPUNKT`). Der ‘wahre’ Schnittpunkt  $v_s$  hat ‘unendlich genaue’ Koordinaten und liegt zwischen den Gitterpunkten. Er wird daher durch den nächstliegenden repräsentierbaren Knoten  $v'_s$  ersetzt (Mitte). Vor der Teilung der sich schneidenden Kanten werden diese sowohl aus Eventfolge  $\mathcal{Q}$  als auch aus der Statusfolge  $\mathcal{T}$  entfernt; anschließend werden die vier Teilkanten  $(v_1, v'_s)$ ,  $(v'_s, v_2)$ ,  $(v_3, v'_s)$ ,  $(v'_s, v_4)$  in  $\mathcal{Q}$  und  $\mathcal{T}$  wieder eingefügt. Da  $v_s$  links von  $(v_1, v_2)$ ,  $v'_s$  dagegen rechts von  $(v_1, v_2)$  liegt, ist die lokale Ordnung

<sup>31</sup>Das kleinstmögliche repräsentierbare Dreieck ist das rechtwinklige Dreieck, dessen Eckpunkte in drei benachbarten Gitterpunkten liegen. Die zweifache Fläche dieses Dreiecks ist also  $2A_{min} = d^2$ . Bei extrem kleinen Dreiecksflächen  $A_D$  ist die Entscheidung, ob drei Punkte kollinear sind, unsicher; wir müssen daher einen Schwellenwert  $\tau$  festlegen, so daß wir auf Kollinearität schließen können, wenn  $|2A_D| < \tau$ . Der Schwellenwert muß groß genug sein, um für alle  $|2A_D| \geq \tau$  eine sichere Entscheidung garantieren zu können. Die Gitterweite  $d$  muß wiederum mindestens so groß sein, daß  $d^2 \geq \tau$ . Zur Wahl der Größe von  $d$  bzw.  $\varepsilon$  vgl. auch Fortune, Milenkovic 1991.

<sup>32</sup>Der Einfachheit halber verzichten wir im folgenden auf die Unterscheidung zwischen einem Modell  $M$  und der Repräsentation  $R$ . Knoten, und Kanten, die sowohl zu  $R$  als auch zu  $R^*$  gehören, werden jetzt einfach mit  $v$  bzw.  $e$  bezeichnet (also ohne Strich); Schnittpunkte in  $R^*$  mit einem Strich gekennzeichnet, z.B.  $v'$ .





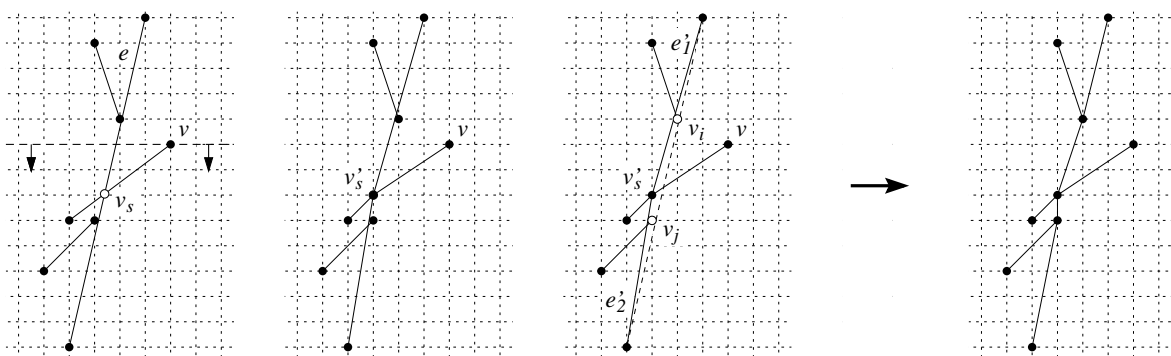
**Abbildung 3.28** Überziehung von Kanten als Folge verzerrter Topologie. Durch die Rundung der Koordinaten des Schnittpunktes  $v_s$  ändert sich während der Bearbeitung des Ereignispunktes  $v_3$  die Reihenfolge von Kanten in der Statusfolge  $\mathcal{T}$ . Bleibt dies unberücksichtigt, dann liegt im nächsten Eventpunkt  $v_s$  ein unentdeckter Schnittpunkt oberhalb der Scanlinie. Um dies zu verhindern, müssen alle Kanten, die zwischen  $v_s$  und  $v'_s$  liegen, zu  $v'_s$  ‘gezogen’ werden.

in  $\mathcal{T}$  jetzt  $(v_1, v'_s) \prec_{\mathcal{T}} (v_1, v_5) \prec_{\mathcal{T}} (v_3, v'_s)$  (Mitte), d.h.  $(v_3, v'_s)$  hat nun einen neuen linken Nachbarn und mit diesem einen Punkt gemeinsam. Allerdings werden  $(v_3, v'_s)$  und  $(v_1, v_5)$  nicht auf Intersektion geprüft, vielmehr wird  $v'_s$  zum neuen Eventpunkt, mit der Folge, daß ein unentdeckter Schnittpunkt *oberhalb* der Scanlinie existiert, der im weiteren Verlauf niemals gefunden werden kann (rechts)!

Diese Verletzung der Invariante könnte nun relativ leicht umgangen werden, indem im aktuellen Eventpunkt  $v_3$  die Kante  $(v_3, v_4)$  bzw. das Teilstück, das nach der ersten Teilung an  $v_3$  gebunden ist, so lange mit der in  $\mathcal{T}$  links benachbarten Kante auf Intersektion geprüft wird, bis kein Schnittpunkt mehr gefunden wird. Wie Abb. 3.29 zeigt, ist eine solche Korrektur in anders gelagerten Fällen völlig unzureichend: Bei der Bearbeitung des aktuellen Eventpunktes  $v$  wird der Schnittpunkt  $v_s$  entdeckt (links) und durch den nächstliegenden repräsentierbaren Knoten  $v'_s$  ersetzt (Mitte links). Dadurch wird nun nicht, wie oben, der Schnittpunkt über eine bestehende Kante ‘gezogen’, sondern die durch die Teilung in  $v'_s$  entstandenen Teilkanten *überziehen* zwei Knoten, so daß sowohl unterhalb als auch *oberhalb* des aktuellen Eventpunktes neue Schnittpunkte entstehen (Mitte links und Mitte rechts). Eine solche Verletzung der Invariante kann nun nicht lokal über eine geschickte Nutzung der Statusfolge gelöst werden, denn der Knoten  $v_i$  und alle an ihn gebundenen Kanten liegen vollständig oberhalb der Scanlinie, gehören daher auch nicht zur Statusfolge und sind im aktuellen Eventpunkt vollkommen unbekannt und nicht ohne weiteres ermittelbar.

### Eine topologisch konsistente Nagelbrettgeometrie

Die einzig sinnvollen Lösungen der Probleme in Abb. 3.28 und 3.29 sind jeweils ganz rechts dargestellt: Das Überziehen von Knoten und Kanten muß verhindert werden, so daß nach der ‘Verschiebung’ des Schnittpunktes  $v_s$  ein Kantenzug entsteht, der über  $v'_s$  verläuft und sich an die bereits existierenden Knoten topologisch konsistent ‘anlegt’. Als analoges Bild für dieses Problem können wir uns ein Nagelbrett vorstellen, auf das ein regelmäßiges Gitter gezeichnet ist. An einigen Gitterpunkten stecken Nägel (die Knoten), die durch elastische Schnüre (die Kanten) verbunden sind. Wir suchen die Stellen, an denen kein Nagel ist und an denen sich zwei oder mehr Schnüre überkreuzen. An jeder dieser Stellen tun wir folgendes: Zunächst bestimmen wir den Punkt des Gitters, der dem Überkreuzungspunkt am nächsten liegt. Steckt auf diesem Gitterpunkt noch kein Nagel, so müssen wir ihn setzen. Anschließend tippen wir mit einem Finger so auf das Brett, daß die sich kreuzenden Schnüre zwischen dem



**Abbildung 3.29** Überziehung von Knoten als Folge verzerrter Topologie. Durch die Rundung der Koordinaten von Schnittpunkten können neue Schnittpunkte während des Scanline-Schwenks sowohl unterhalb als auch oberhalb der aktuellen Scanlinie entstehen. Dieses ‘Überziehen’ von Knoten kann verhindert werden, wenn die Knoten bekannt sind, die ‘überzogen’ werden können; dann nämlich kann eine Kante wie eine Schnur auf einem Nagelbrett ‘um diese Knoten gelegt’ werden.

Finger und diesem ‘Zielnagel’ liegen.<sup>33</sup> Wir ziehen dann unseren Finger in Richtung dieses Nagels, wobei wir ständig das Brett berühren, und achten darauf, daß wir keine Schnur unter unserem Finger hindurchlassen und daß sich auch keine Schnüre von anderen Nägeln lösen. Wenn wir schließlich den Zielnagel berühren, liegen genau jene Schnüre an unserem Finger, die wir über diesen Nagel legen müssen.<sup>34</sup> Wenn wir vorsichtig genug waren, haben sich die gezogenen Schnüre zudem automatisch um die richtigen Nägel gelegt, die ihnen während des Ziehens gewissermaßen im Weg waren.<sup>35</sup>

Die Frage ist nun, wie wir die Idee des analogen Nagelbretts in ein effektives digitales Verfahren umsetzen können. Gibt es eine Möglichkeit, feste Nägel und elastische Schnüre in den Scanline-Algorithmus zu integrieren? Das Problem ist ja, daß wir während des Scanline-Verfahrens immer nur einen kleinen Ausschnitt unseres Nagelbretts ‘sehen’ können, daß unsere Nägel nicht physisch, sondern über zwei Zahlen auf einem Brett ‘verankert’ sind, das wir nicht berühren können, und daß unsere Schnüre weder elastisch sind noch die Fähigkeit aufweisen, automatisch an (physisch nicht vorhandenen) Nägeln ‘hängenzubleiben’. Im Prinzip ist die Lösung dennoch sehr einfach, denn wie wir sehen, können zwei verschiedene Fälle (die natürlich ggf. in Kombination auftreten können) unterschieden werden:

- Ein Schnittpunkt wird über eine bestehende Kante  $e$  ‘gezogen’: Liegt der ‘wahre’ Schnittpunkt  $v_s$  links (rechts) von  $e$ , dann liegt der repräsentierbare Schnittpunkt  $v'_s$  rechts (links) von  $e$  oder auf  $e$ .
- Eine Kante wird über einen bestehenden und repräsentierbaren Knoten  $v_r$  ‘gezogen’: Liegt  $v_r$  links (rechts) von der ‘wahren’ Kante  $e$ , dann liegt  $v_r$  rechts (links) von der repräsentierbaren Kante  $e'$  oder auf  $e'$ .

Beide Fälle lassen sich also auf den Wechsel einer lokalen Ordnung reduzieren: Gilt für eine Kante  $e$  und einen Knoten  $v$  in  $v$  die Ordnung  $e \prec_{\mathcal{T}} v$ , so gilt nach einer Verletzung der Invariante die Ordnung  $v \preceq_{\mathcal{T}} e$  (und umgekehrt). Nun kennen wir die lokale Ordnung  $\prec_{\mathcal{T}}$  nur für den aktuellen Eventpunkt, d.h. als Ordnung der Statusfolge  $\mathcal{T}$ . Wenn wir dagegen für die Kante  $e_1$  in Abb. 3.29 (Mitte rechts)

<sup>33</sup>Die Startposition des Fingers liegt in Abb. 3.28 senkrecht oberhalb von  $v_s$ , in Abb. 3.29 rechts von  $v_s$ .

<sup>34</sup>In Abb. 3.28 sind dies die Kanten  $(v_1, v_2)$  und  $(v_1, v_5)$ .

<sup>35</sup>In Abb. 3.29 sind dies die Knoten  $v_i$  und  $v_j$ .

wissen, daß der Knoten  $v_i$  eigentlich (aus unserer Sicht) rechts von ihr liegt, können wir direkt nach der Teilung der Kante  $e$ , d.h. vor dem Wiedereinfügen der (Teil-)Kante  $e_1$  in die Event- und Statusfolge, überprüfen, ob  $v_i$  ebenfalls links von  $e_1$  ist. Ist das, wie hier, nicht der Fall, müssen wir die Kante  $e_1$  erneut teilen, und zwar in  $v_i$ . Führen wir die entsprechende Überprüfung auch für  $e_2$  und  $v_j$  durch, erhalten wir das Ergebnis in Abb. 3.29 (rechts). Ähnlich kann auch auf die Situation in Abb. 3.28 reagiert werden. Hier muß bekannt sein, auf welcher Seite von  $(v_1, v_5)$  der ‘wahre’ Schnittpunkt  $v_s$  liegt, um vergleichen zu können, ob der repräsentierbare Schnittpunkt  $v'_s$  auf derselben Seite liegt. Ist dies, so wie hier, nicht der Fall, muß die Kante  $(v_1, v_5)$  ebenfalls in  $v'_s$  geteilt werden.

Offensichtlich können wir das Problem lösen, indem wir uns einzig auf symbolische Informationen verlassen, die uns die korrekte Ordnung von Knoten und Kanten angeben. Zu Beginn des Verfahrens ist die Ordnung aller Kanten, so wie sie in der doppelt-verketteten Kantenliste  $\mathcal{D}$  vorliegen, ‘relativ’ korrekt, denn  $\mathcal{D}$  enthält die Repräsentation des  $\varepsilon$ -nahen Modells aller ‘wahren’ Kanten. Kritisch sind also nur neue Knoten, die während des Scanline-Schwenks gefunden werden, d.h. wir müssen nur dann auf Inkonsistenzen überprüfen, wenn ein Schnittpunkt gefunden wird. Die Frage ist dann aber, welche Knoten und Kanten auf die Einhaltung ihrer jeweiligen lokalen Ordnung geprüft werden müssen.

### Eingrenzung des Überprüfungsbereichs

Wir können zunächst festhalten, daß alle neuen Schnittpunkte (im folgenden *derivierte Schnittpunkte*), die durch die Verschiebung eines ‘wahren’ Schnittpunktes  $v_s$  auf den repräsentierbaren Schnittpunkt  $v'_s$  entstehen können, dann unproblematisch sind, wenn sie unterhalb von  $v'_s$  auftreten und  $v'_s$  unterhalb des aktuellen Eventpunktes liegt. Genau das behauptet auch der folgende Hilfssatz (vgl. Abb. 3.28 und 3.29):

**Hilfssatz 3.8.** *Jeder derivierte Schnittpunkt  $v_d$ , der während des Scanline-Verfahrens bei der Bearbeitung des aktuellen Eventpunktes  $v_a$  durch die Teilung einer Kante im repräsentierbaren Schnittpunkt  $v'_s$  entsteht, wird im weiteren Verlauf des Scanline-Schwenks gefunden, wenn  $v_a \prec_Q v'_s \prec_Q v_d$ .*

*Beweis.* Da  $v_a \prec_Q v'_s$ , liegt jeder Schnittpunkt, der neu in die Eventfolge  $Q$  eingefügt wird, unterhalb der aktuellen Scanlinie (Schnittpunkte, die oberhalb oder links von  $v_a$  gefunden werden, bleiben in ERZEUGESCHNITTPUNKTEVENT unberücksichtigt). Daher wird  $v'_s$  nach der Bearbeitung von  $v_a$  zum aktuellen Eventpunkt. Sei  $e = (v_u, v_l)$  die Kante, die in  $v'_s$  geteilt wird, sei  $v_u$  der obere Endpunkt von  $e$  und seien  $e'_i$  und  $e'_j$  die beiden Kanten, die sich nach der Teilung von  $e$  in  $v_d$  schneiden. Wenn  $v'_s \prec_Q v_d$ , dann liegt der derivierte Schnittpunkt  $v_d$  nach der Teilung von  $e$  auf der Kante  $e'_i = (v'_s, v_l)$ , wobei  $v'_s$  der obere Endpunkt dieser Kante ist. Da  $v_a \prec_Q v'_s$ , ist  $e'_i$  nach der Bearbeitung des aktuellen Eventpunktes  $v_a$  nicht in der Statusfolge  $\mathcal{T}$  enthalten. Der Schnittpunkt  $v_d$  kann daher frühestens dann gefunden werden, wenn  $v'_s$  der aktuelle Eventpunkt ist und  $e'_i$  in  $\mathcal{T}$  eingefügt wurde. Daher gibt es einen Eventpunkt  $v$ ,  $v'_s \preceq_Q v \prec_Q v_d$ , an dem  $e'_i$  und  $e'_j$  in  $\mathcal{T}$  direkt benachbart sind und auf Intersektion geprüft werden.  $\square$

Ebenso wird jeder derivierte Schnittpunkt im weiteren Verlauf des Scanline-Schwenks gefunden, wenn er auf einer Kante liegt, die vollständig unterhalb der aktuellen Scanlinie liegt:

**Hilfssatz 3.9.** *Jeder derivierte Schnittpunkt  $v_d$ , der während des Scanlineverfahrens bei der Bearbeitung des aktuellen Eventpunktes  $v_a$  durch die Teilung einer Kante im repräsentierbaren Schnittpunkt  $v'_s$  entsteht, wird im weiteren Verlauf des Scanline-Schwenks gefunden, wenn wenigstens eine Kante  $e = (v_u, v_l)$  existiert, die  $v_d$  enthält und vollständig unterhalb der aktuellen Scanlinie liegt, d.h. wenn  $v_d \in e$ ,  $v_a \prec_Q v_u$  und  $v_a \prec_Q v_l$ .*

*Beweis.* Wir können ganz analog zum obigen Beweis argumentieren: Sei  $v_u$  der obere Endpunkt von  $e$ , d.h.  $v_u \prec_Q v_l$ , und sei  $e_d$  die Kante, die sich mit  $e$  in  $v_d$  schneidet. Der Schnittpunkt  $v_d$  kann frühestens dann gefunden werden, wenn  $v_u$  der aktuelle Eventpunkt ist und  $e$  in  $\mathcal{T}$  eingefügt wurde. Daher gibt es einen Eventpunkt  $v$ ,  $v_u \preceq_Q v \prec_Q v_d$ , an dem  $e$  und  $e_d$  in  $\mathcal{T}$  direkt benachbart sind und auf Intersektion geprüft werden.  $\square$

Ob durch die Projektion eines Schnittpunktes  $v_s$  auf ein diskretes Raster derivierte Schnittpunkte entstehen, muß also nur oberhalb des repräsentierbaren Schnittpunktes  $v'_s$  überprüft werden und auch nur für solche Kanten, die die aktuelle Scanlinie schneiden oder vollständig oberhalb der Scanlinie liegen. Die Überprüfung erfolgt anhand der lokalen Ordnungen von Knoten und Kanten vor und nach der Teilung der sich schneidenden Kanten in  $v'_s$ : Ist die Ordnung identisch, liegt kein derivierter Schnittpunkt vor, ist sie dagegen verschieden, muß eine erneute Teilung einer (Teil-)Kante vorgenommen werden. Für jede zu teilende Kante muß daher bekannt sein, welche Knoten in ihrer Nähe links und rechts von ihr liegen (Abb. 3.29). Für alle anderen Kanten, die in der Nähe von  $v_s$  liegen, muß zudem bekannt sein, auf welcher Seite  $v_s$  liegt, um diese Lage mit der Lage von  $v'_s$  nach der Teilung vergleichen zu können (Abb. 3.28).

Wie nah müssen sich Knoten und Kanten sein, damit sich überhaupt derivierte Schnittpunkte ergeben können, und woher kennen wir sie? Sei  $d(e, v)$  die Distanz (d.h. die Länge der Lotrechten) zwischen einem Knoten  $v$  und einer Kante  $e$ . Liegt eine Kante  $e$  zwischen einem Schnittpunkt  $v_s$  und seiner Repräsentation  $v'_s$  (Abb. 3.28), ist natürlich  $d(v_s, e) \leq eps$ . Wenn sich die relative Lage eines Knotens  $v_i$  zu einer Kante  $e$  von der relativen Lage zu einer der beiden Teilstücke von  $e$  unterscheidet (Abb. 3.29), ist ebenfalls  $d(v_i, e) \leq eps$ . Dennoch reicht es nicht aus, wenn wir uns für jede Kante nur diejenigen Knoten merken, deren Distanz kleiner als  $\varepsilon$  ist. Der Grund liegt darin, daß wir ein effektives Verfahren benötigen, um die zu überprüfenden Knoten/Kantenpaare zu finden. Das können wir nur erreichen, wenn wir die links und rechts liegenden Knoten zusammen mit den entsprechenden Kanten speichern, so daß darauf jederzeit direkt zurückgegriffen werden kann. Betrachten wir nun Abb. 3.27, wird deutlich, daß es Situationen geben kann, in denen Kanten(stücke) im Laufe des Scanline-Schwenks immer mehr in die Richtung von ursprünglich weiter entfernten Knoten 'verboten' werden. Für einige Kanten kann also während des Verfahrens die Kenntnis der relativen Lage eines Knotens bedeutsam werden, der zu Beginn des Verfahrens in einer Distanz lag, die deutlich größer als  $\varepsilon$  ist.

Natürlich müssen wir uns deshalb nicht gleich alle linken und rechten Knoten merken. Es genügt, wenn wir uns auf die *Bounding Box* beschränken, also das kleinste achsenparallele Rechteck, mit dem eine Kante umschlossen werden kann. Mit dem folgenden Hilfssatz wird behauptet, daß die Ausdehnung jedes Halbkantenzugs, unabhängig davon, wie oft er während des Scanline-Verfahrens geteilt wird, immer durch seine ursprüngliche Bounding Box beschränkt ist:

**Hilfssatz 3.10.** *Sei  $e = (v_u, v_l)$  eine Kante,  $v_u$  und  $v_l$  seien mit fester Präzision repräsentierbar. Werden  $e$  bzw. die aus  $e$  entstandenen Teilstücke  $e_i$ ,  $1 \leq i \leq k + 1$ , in  $k$  beliebigen, auf  $e$  bzw.  $e_i$  liegenden Punkten geteilt, und wird jeder Teilungspunkt durch den nächstliegenden repräsentierbaren Punkt ersetzt, dann liegt der entstehende Kantenzug, der  $v_u$  und  $v_l$  nach Abschluß des Verfahrens verbindet, niemals außerhalb eines Rechtecks  $R$ , dessen gegenüberliegende Ecken in  $(\min(v_l.x, v_u.x), \min(v_l.y, v_u.y))$  und  $(\max(v_l.x, v_u.x), \max(v_l.y, v_u.y))$  liegen.*

*Beweis.* Sei  $p$  der Punkt, in dem die Kante  $e$  geteilt werden soll, und sei  $p'$  die Repräsentation von  $p$ . Da  $p \in e$ , ist  $\min(v_l.x, v_u.x) \leq p.x \leq \max(v_l.x, v_u.x)$  und  $\min(v_l.y, v_u.y) \leq p.y \leq \max(v_l.y, v_u.y)$ . Da die Koordinaten von  $v_u$  und  $v_l$  repräsentierbar sind, gilt für  $p'$  entsprechend  $\min(v_l.x, v_u.x) \leq p'.x \leq \max(v_l.x, v_u.x)$  und  $\min(v_l.y, v_u.y) \leq p'.y \leq \max(v_l.y, v_u.y)$ , d.h.  $p'$  liegt in  $R$ . Da  $v_l$  und

$v_u$  auf zwei gegenüberliegenden Ecken von  $R$  liegen, liegt auch jede beliebige Kante  $e_i = (v_l, p')$  und  $e_j = (v_u, p')$  in  $R$  oder auf dem Rand von  $R$ , jedoch niemals außerhalb von  $R$ . Daher liegen auch die kleinsten achsenparallelen Rechtecke, mit denen  $e_i$  bzw.  $e_j$  umschlossen werden können, innerhalb von  $R$ . Dasselbe gilt dann auch für die Kanten, die durch die erneute Teilung von  $e_i$  bzw.  $e_j$  entstehen. Daher liegt der entstehende Kantenzug, der  $v_u$  und  $v_l$  nach Abschluß des Verfahrens verbindet, niemals außerhalb von  $R$ .  $\square$

### Überprüfung lokaler Ordnungen

Um die zu überprüfenden Knoten/Kanten-Ordnungen effektiv bearbeiten zu können, führen wir zusammen mit jeder Kante  $e$  in der doppelt-verketteten Kantenliste  $\mathcal{D}$  drei Listen, die die folgenden Informationen enthalten: Die Liste  $V_l(e)$  enthält alle Knoten, die innerhalb der Bounding-Box von  $e$  links von  $e$  liegen, die Liste  $V_r(e)$  enthält alle Knoten, die innerhalb der Bounding-Box von  $e$  rechts von  $e$  liegen, und die Liste  $V_c(e)$  enthält alle Knoten, die auf  $e$  liegen. Bei der Teilung einer Kante  $e$  werden, analog zu den Flächeninformationen, auch die Knotenlisten  $V_l(e)$ ,  $V_r(e)$  und  $V_c(e)$  kopiert, und zwar so, daß die Knotenlisten jedes der entstandenen Teilstücke  $e'$  nur die Knoten enthält, die innerhalb der Bounding-Box von  $e'$  liegen. Bei der Auflösung einer Überlappung werden die Knotenlisten entsprechend vereinigt.

Betrachten wir zunächst Abb. 3.28. Zwei Kanten  $(v_1, v_2)$  und  $(v_3, v_4)$  schneiden sich in  $v_s$ , und die Kante  $e = (v_1, v_5)$  liegt zwischen dem Schnittpunkt  $v_s$  und seiner Repräsentation  $v'_s$ . Zu bestimmen ist die Menge  $E$  der Kanten, die sich nicht in  $v_s$  schneiden und innerhalb des kleinstmöglichen Rechtecks liegen, dessen Grenze durch die Knoten  $v_1$  (dem oberen Endpunkt von  $(v_1, v_2)$ ),  $v_3$  (dem oberen Endpunkt von  $(v_3, v_4)$ ) und  $v'_s$  verläuft. Diese Menge enthält hier genau eine Kante, nämlich  $(v_1, v_5)$ . Wir bestimmen nun die relative Lage des 'wahren' Schnittpunktes  $v_s$  zu jeder Kante  $e_i \in K$ ; für die Kante  $(v_1, v_5)$  erhalten wir als Ergebnis, daß  $v_s$  links<sup>36</sup> von ihr liegt. Zusammen mit jeder Kante  $e_i \in K$  speichern wir  $v'_s$  in der Liste  $V_l(e_i)$ , die ausschließlich links benachbarte Knoten enthält. Jetzt teilen wir die sich schneidenden Kanten in  $v'_s$  und bestimmen die Menge  $E$  erneut.  $E$  enthält jetzt statt der Kante  $(v_1, v_2)$  das obere Teilstück  $(v_1, v'_s)$ . Ist die relative Lage von  $v'_s$  zu den Kanten in  $E$  nach der Teilung identisch mit der relativen Lage von  $v_s$ , dann liegen keine derivierten Schnittpunkte oberhalb von  $v'_s$  vor. In  $v_s$  gilt in unserem Beispiel  $(v_1, v_5) \prec_{\mathcal{T}} v_s$ , in  $v'_s$  dagegen  $v'_s \prec_{\mathcal{T}} (v_1, v_5)$ . Daher wissen wir, daß wir bei der 'Verschiebung' von  $v_s$  nach  $v'_s$  die Kante  $(v_1, v_5)$  'überzogen' haben und sie ebenfalls in  $v'_s$  teilen müssen. Um dies zu überprüfen, testen wir für jede Kante  $e_i$  aus  $E$  direkt nach ihrer Teilung, ob es (1.) einen Knoten gibt, der in  $V_l(e_i)$  gespeichert ist, aber nicht links von  $e_i$  liegt, ob es (2.) einen Knoten gibt, der in  $V_r(e_i)$  gespeichert ist, aber nicht rechts von  $e_i$  liegt oder ob es (3.) einen Knoten gibt, der in  $V_c(e_i)$  gespeichert ist. Sobald wir einen Knoten gefunden haben, für den eines der drei Kriterien zutrifft, müssen wir die Kante  $e_i$  in diesem Knoten teilen und direkt nach dieser Teilung diese Prüfung für jedes der entstandenen Teilstücke erneut durchführen. In Fall von Abb. 3.28 würden wir bei der Überprüfung der Kante  $e = (v_1, v_5)$  feststellen, daß der Knoten  $v'_s$  in der Liste  $V_l(e)$  gespeichert ist, aber rechts von  $e$  liegt. Daher muß  $(v_1, v_5)$  auch in  $v'_s$  geteilt werden.

Betrachten wir nun Abb. 3.29. Wiederum schneiden sich zwei Kanten  $(v_1, v_2)$  und  $(v_3, v_4)$  in  $v_s$ . Vor der Teilung der beiden sich schneidenden Kanten liegt  $v_6$  relativ gesehen rechts von der Kante  $(v_1, v_2)$ , nach der Teilung in  $v'_s$  dagegen links von dem entstandenen Kantenzug von  $v_1$  über  $v'_s$  nach  $v_2$ . Zu bestimmen ist die Menge  $K$  der Knoten, die von den Kanten  $e_1 = (v_1, v_2)$  und  $e_2 = (v_3, v_4)$  durch die Teilung in  $v'_s$  überzogen werden können. Diese Knoten liegen alle innerhalb des kleinstmöglichen Rechtecks, dessen Grenze durch die Knoten  $v_1$  (dem oberen Endpunkt von  $(v_1, v_2)$ ),  $v_3$  (dem oberen Endpunkt von  $(v_3, v_4)$ ) und  $v'_s$  verläuft.  $K$  enthält daher die Knoten  $v_5$  und  $v_6$ . Diese Knoten müssen

<sup>36</sup>Man beachte die Richtung der Kante von  $v_1$  nach  $v_5$ !

in die Knotenlisten der beiden sich schneidenden Kanten  $(v_1, v_2)$  und  $(v_3, v_4)$  eingetragen sein; für die Kante  $e_1$  ist  $V_r(e_1) = \{v_5, v_6\}$ ,  $V_l(e_1)$  und  $V_c(e_1)$  sind leer. Entsprechend ist nach der Teilung mit dem oberen Teilstück  $e_{11} = (v_1, v'_s)$  die Knotenliste  $V_r(e_{11}) = \{v_5, v_6\}$  gespeichert,  $V_l(e_{11})$  und  $V_c(e_{11})$  sind wiederum leer. Bei der Überprüfung von  $e_{11}$  wird daher der Knoten  $v_6$  gefunden, der in  $V_r(e_{11})$  gespeichert ist, jedoch links von  $e_{11}$  liegt. Daher wird  $e_{11} = (v_1, v'_s)$  erneut in  $v_6$  geteilt. Der derivierte Schnittpunkt zwischen den Kanten  $(v_1, v_2)$  und  $(v_7, v_8)$  bleibt unentdeckt, wird jedoch nach Satz 3.8 im weiteren Verlauf des Scanline-Schwenks gefunden.

In beiden Fällen müssen wir unsichere Entscheidungen in Kauf nehmen; immer, wenn nicht wirklich sicher ist, ob ein ‘wahrer’ Schnittpunkt links oder rechts von einer Kante liegt, nehmen wir an, er läge auf der Kante. Deshalb ist es von zentraler Bedeutung, daß wir bei der Überprüfung einer Kante  $e$  die Liste  $V_c(e)$  führen und  $e$  immer auch dann erneut teilen, wenn  $V_c(e)$  ein Element enthält.

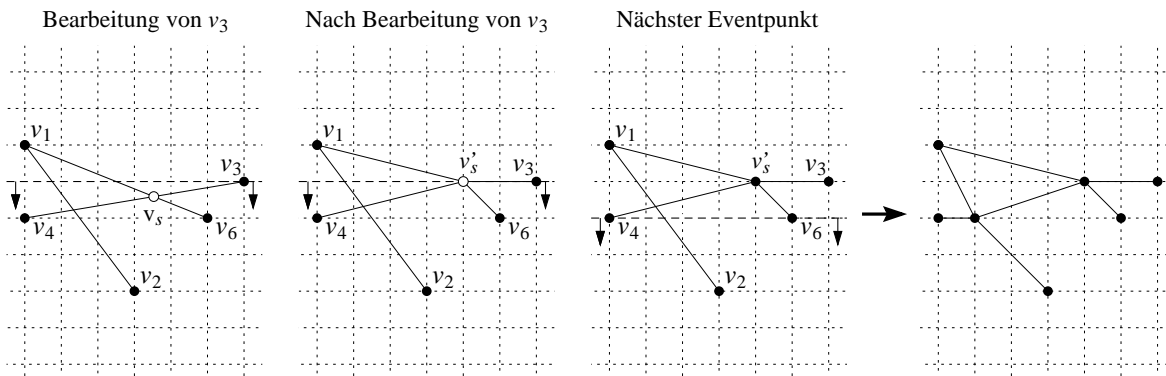
### Bestimmung der überprüften Knoten und Kanten

Beim Überziehen eines Schnittpunktes über eine oder mehrere Kanten (Abb. 3.28) müssen alle Kanten bekannt sein, die innerhalb des Rechtecks zwischen den oberen beiden Endpunkten der sich schneidenden Kanten und dem neuen Schnittpunkt liegen. Da die sich schneidenden Kanten auch die Scanlinie schneiden und der Schnittpunkt unterhalb der Scanlinie liegt, liegen die möglicherweise überzogenen Kanten entweder auf oder unterhalb der Scanlinie. Nach Hilfssatz 3.9 können wir die Kanten, die vollständig unterhalb der Scanlinie liegen, ignorieren. Daher werden die gesuchten Kanten alle von der Scanlinie geschnitten und sind in der Statusfolge  $\mathcal{T}$  gespeichert.

Beim Überziehen einer Kante über einen oder mehrere Knoten (Abb. 3.29) müssen alle Knoten bekannt sein, die innerhalb des Rechtecks zwischen den oberen beiden Endpunkten der sich schneidenden Kanten und dem neuen Schnittpunkt liegen. Bei der Bearbeitung eines Eventpunktes  $v$  können zunächst die zu  $v$  in der Statusliste  $\mathcal{T}$  direkt benachbarten Kanten bestimmt werden (in Zeile 7 in BEARBEITEEVENTPUNKT), so daß  $v$  in die entsprechenden Knotenlisten dieser Kanten eingetragen werden kann. Auf diese Weise spiegeln die Knotenlisten jeder Kante die korrekten Verhältnisse jeweils oberhalb der aktuellen Scanlinie wider. Alle anderen Knoten, die gegebenenfalls überprüft werden müssen, liegen auf oder unterhalb der Scanlinie und nach Hilfssatz 3.8 oberhalb des neuen Schnittpunktes. Nach Hilfssatz 3.9 sind diese Knoten ausschließlich Endpunkte von Kanten, die von der aktuellen Scanlinie geschnitten werden, d.h. sie sind in der Statusfolge  $\mathcal{T}$  gespeichert. Ohne Beweis führen wir daher an:

**Hilfssatz 3.11.** *Wenn oberhalb der aktuellen Scanlinie für jede Kante die direkt links und rechts benachbarten Knoten bekannt sind, dann sind alle weiteren, zur Prüfung auf derivierte Schnittpunkte notwendigen Knoten und Kanten in der Statusfolge  $\mathcal{T}$  gespeichert.*

Leider können wir die Statusfolge  $\mathcal{T}$  nicht direkt nach einer Menge von Kanten ‘befragen’, die innerhalb eines bestimmten Rechtecks liegen. Aber da wir nach Hilfssatz 3.11 wissen, daß die gesuchten Kanten bzw. deren Endpunkte in  $\mathcal{T}$  sind, können wir zusätzlich zu  $\mathcal{T}$  eine weitere Datenstruktur führen, die eine solche Abfrage erlaubt. Diese Datenstruktur nennen wir *Intervallfolge*, bezeichnen sie mit  $\mathcal{I}$  und implementieren sie als dynamischen Intervallbaum (vgl. Ottmann, Widmayer 1993, S. 500ff.). In  $\mathcal{I}$  speichern wir für jede Kante  $(v_i, v_j) \in \mathcal{T}$  das horizontale Intervall  $[\min(v_i.x, v_j.x), \max(v_i.x, v_j.x)]$ . Jedesmal, wenn wir eine Kante in  $\mathcal{T}$  einfügen, fügen wir dieselbe Kante auch in  $\mathcal{I}$  ein, und jedesmal, wenn wir eine Kante aus  $\mathcal{T}$  löschen, löschen wir sie auch aus  $\mathcal{I}$ . Wenn sich zwei Kanten, deren obere Endpunkte  $v_i$  und  $v_j$  sind, im Punkt  $v_s$  schneiden, dann können wir die Menge der zur Konsistenzprüfung notwendigen Knoten und Kanten bestimmen, indem wir in  $\mathcal{I}$  nach allen Kanten suchen, deren  $x$ -Intervalle sich mit dem Intervall  $[\min(v_i.x, v_j.x, v_s.x), \max(v_i.x, v_j.x, v_s.x)]$  überschneiden. Ein-



**Abbildung 3.30** Versetzung von Eventpunkten als Folge verzerrter Topologie. Ein ‘wahrer’ Schnittpunkt  $v_s$  liegt immer unterhalb der Scanlinie, d.h. unterhalb des aktuellen Eventpunktes ( $v_3$ ). Dies gilt jedoch nicht immer für dessen Repräsentation  $v'_s$ . Liegt  $v'_s$  in der Eventfolge  $\mathcal{Q}$  vor dem aktuellen Eventpunkt, dann muß in  $\mathcal{Q}$  bis  $v'_s$  zurückgeschritten werden.

fügen und Löschen einer Kante in  $\mathcal{I}$  benötigt  $O(\log n)$  Zeit; die für die Suche benötigte Zeit liegt in  $O(\log n + k)$ , wobei  $k$  die Anzahl der gefundenen Kanten ist.

### 3.5.3.2 Globale Verletzungen der Invariante

Eine weitere Verletzung der Invariante tritt auf, wenn ein Schnittpunkt  $v_s$  gefunden wird, dessen  $\varepsilon$ -nahe Repräsentation  $v'_s$  vor dem aktuellen Eventpunkt liegt (Abb. 3.30). Solche *Versetzungen* treten nur dann auf, wenn mindestens eine der beiden sich schneidenden Kanten annähernd horizontal ist, so daß ein Schnittpunkt auf die Scanlinie gewissermaßen ‘hochgerundet’ wird. Ohne die Projektion von  $v_s$  auf  $v'_s$  würde nach der Bearbeitung von  $v_3$  der Schnittpunkt  $v_s$  zum aktuellen Eventpunkt, bei dessen Bearbeitung die beiden Kanten  $(v_1, v_6)$  und  $(v_3, v_4)$  auf Intersektion geprüft würden. Wird  $v_s$  dagegen durch  $v'_s$  ersetzt, ist der nächste Eventpunkt nach  $v_3$  der Knoten  $v_4$ , so daß oberhalb der Scanlinie ein unentdeckter Schnittpunkt existiert (Mitte rechts).

Wie Abb. 3.30 zeigt, können bei Versetzungen, im Gegensatz zu Überziehungen, von vorneherein existierende Schnittpunkte unentdeckt bleiben, d.h. Schnittpunkte, die nicht erst als Derivat einer verzerrten Topologie entstanden sind. Wir haben daher keine Möglichkeit, solche Fehler mit Hilfe einer Nagelbrettgeometrie zu umgehen. Mit anderen Worten: Wir können uns nicht auf die Prüfung lokaler Ordnungen verlassen, da keine lokalen Ordnungen verletzt werden. Verletzt wird vielmehr die globale Ordnung der Eventfolge  $\mathcal{Q}$ , nach der jeder Schnittpunkt immer unterhalb des aktuellen Eventpunktes liegt und daher im weiteren Verlauf des Scanline-Schwenks immer auch selber zum aktuellen Eventpunkt wird.

Die Lösung des Problems in Abb. 3.30 ist wiederum ganz rechts dargestellt. Um den Schnittpunkt der beiden Kanten  $(v_1, v_6)$  und  $(v'_s, v_4)$  zu finden, müssen wir  $v'_s$  zum aktuellen Eventpunkt machen. Wir dürfen also die Scanlinie nach der Bearbeitung des Eventpunktes  $v_3$  nicht zum nächsten Eventpunkt in der Eventfolge  $\mathcal{Q}$  versetzen (denn das ist  $v_4$ , und  $v_4$  liegt bereits unterhalb des Schnittpunktes), sondern wir müssen in  $\mathcal{Q}$  zurückschreiten, bis wir den versetzten Schnittpunkt  $v'_s$  erreichen und von da aus den Scanline-Schwenk fortsetzen. Dabei müssen wir sehr sorgfältig darauf achten, daß auch die Statusfolge  $\mathcal{T}$  entsprechend aktualisiert wird. Im Falle einer Versetzung wird also mindestens ein Eventpunkt mehrfach bearbeitet: Genau ein Eventpunkt wird mehrfach bearbeitet, wenn  $v'_s$  gleich dem aktuellen Eventpunkt ist oder wenn, wie in Abb. 3.30, zwischen  $v'_s$  und dem aktuellen Eventpunkt kein weiterer Eventpunkt liegt. Es können natürlich auch Fälle auftreten, in denen mehrere Event-

punkte mehrfach abgearbeitet werden müssen. Die zur Auflösung von Versetzungen benötigte Zeit ist daher proportional zur Anzahl der Eventpunkte, die mehrfach bearbeitet werden.

### 3.5.4 Handhabung von (Fast-)Überlappungen

Zu Beginn des Kapitels 3.5.1 wurden mehrere Operationen genannt, für die unsichere Entscheidungen aufgrund unpräziser Arithmetik von Bedeutung sind: die Bestimmung von Schnittpunkten sowie Such-, Einfüge- und Löschooperationen in der Eventfolge, der Statusfolge und der zyklischen Ordnung aller an einen Knoten gebundenen Kanten. Die Tatsache, daß bei der Bestimmung eines Schnittpunktes nach (Gleichung 3.3) Rechenfehler auftreten, bedeutete zunächst, daß die (lokale) relative Lage eines berechneten Schnittpunktes zu seinen benachbarten Kanten unsicher ist, d.h. in manchen Fällen nicht mit der Lage des ‘wahren’ Schnittpunktes übereinstimmt. Unabhängig davon, ob mit berechneten Schnittpunktkoordinaten die Topologie korrekt wiedergegeben wird, treten zusätzliche Unsicherheiten bei der Überprüfung der relativen Lage nach (Gleichung 3.4) auf; bei extrem kleinen Dreiecksflächen treten bei der Berechnung häufig Vorzeichenfehler auf, die nach (3.5) zu ungültigen Schlüssen führen. Durch die Projektion von ‘wahren’ Schnittpunkten auf repräsentierbare Knoten kann schließlich die Invariante verletzt werden, nach der alle Einträge in der doppelt-verketteten Kantenliste  $\mathcal{D}$  links und oberhalb des aktuellen Eventpunktes korrekt sind.

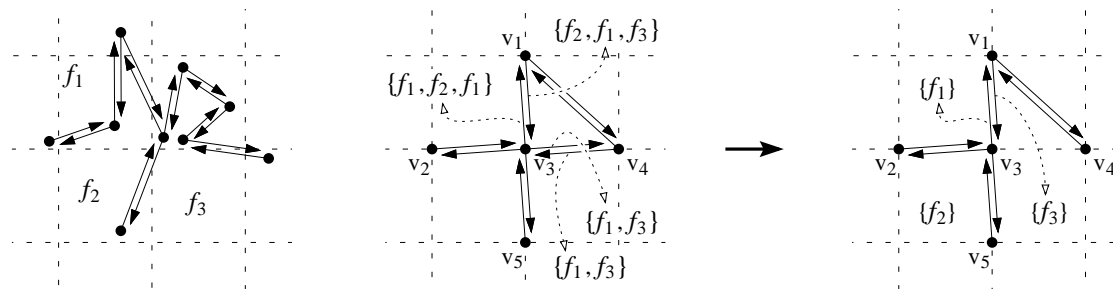
Die notwendigen Korrekturen (Kap. 3.5.3) führen ggf. zu Änderungen in  $\mathcal{D}$  auch oberhalb des aktuellen Eventpunktes. Dabei ist auch nicht ausgeschlossen, daß wir uns neue, gewissermaßen derivierte (Fast-)Überlappungen einhandeln. Erinnern wir uns, daß für uns eine Überlappung immer dann vorliegt, wenn zwei verschiedene Kanten in der Statusfolge  $\mathcal{T}$  nicht unterscheidbar sind. Treten nun (Fast-)Überlappungen oberhalb des aktuellen Eventpunktes auf, besteht keine Notwendigkeit, die betroffenen Kanten in  $\mathcal{T}$  einzufügen. In diesen Fällen bliebe die Überlappung unentdeckt.

Benötigt wird daher ein Verfahren, mit dem wir auch solche (Fast-)Überlappungen finden, die durch die Korrektur von lokalen Verletzungen der Invariante oberhalb der Scanlinie auftreten. Dabei machen wir uns zunutze, daß die Aktualisierung der *Vorgänger*(-) und *Nachfolger*(-)-Zeiger der an einen Eventpunkt gebundenen Kanten die Kenntnis der zyklischen Ordnung dieser Kanten voraussetzt (Kap. 3.2.3.1). Statt für jeden Eventpunkt  $v$  nur diejenigen Kanten zu speichern, deren oberer Endpunkt in  $v$  liegt, können wir ohne zusätzlichen Zeitaufwand auch *alle* Kanten, die an  $v$  gebunden sind, mit  $v$  speichern, und zwar in ihrer zyklischen Ordnung um  $v$ . Zwei Kanten, die sich (fast) überlappen, stehen in der zyklischen Ordnung um  $v$  an derselben Stelle und sind nicht unterscheidbar. Auf diese Weise können Überlappungen sowohl über die Statusfolge  $\mathcal{T}$  als auch an jedem beliebigen Knoten in der Eventfolge  $\mathcal{Q}$  entdeckt werden. Auf zwei verschiedene, nicht unterscheidbare Kanten reagieren wir dann, wie in Kap. 3.4 dargestellt, unmittelbar mit der Auflösung der Überlappung.

### 3.5.5 Aktualisierung der Flächenlabel

Durch eine topologische Verzerrung von Kanten treten bereits bei der Initialisierung der Eventfolge  $\mathcal{Q}$  und der doppelt-verketteten Kantenliste  $\mathcal{D}$  — und später auch während des Scanline-Schwenks — sehr häufig Überlappungen auf. Im Unterschied zu den bislang besprochenen Überlappungsfällen in Kap. 3.4 können sich dabei auch zwei Kanten überlappen, die zu ein und derselben Flächengrenze gehören. Abb. 3.31 zeigt einen solchen Fall. Dargestellt ist ein Ausschnitt einer doppelt-verketteten Kantenliste, so wie sie vor der Initialisierung von  $\mathcal{Q}$  und  $\mathcal{D}$ , d.h. vor der Rundung der Knoten auf die Gitterpunkte, vorliegt (links). Während der Initialisierung von  $\mathcal{Q}$  und  $\mathcal{D}$  müssen zahlreiche Überlappungen aufgelöst werden. Die Kante  $(v_3, v_4)$  resultiert z.B. aus einer zweifachen und die Kante  $(v_1, v_3)$  aus einer dreifachen Überlappung, so daß insgesamt nur noch fünf verschiedene Kanten in  $\mathcal{D}$  existieren (Mitte).





**Abbildung 3.31** Die Bestimmung der Flächenlabel bei konsistent verzerrter Topologie. Durch die topologische Verzerrung von Kanten entstehen Überlappungen. Um für diese zu entscheiden, welche Kanten welche Flächen begrenzen, müssen die ursprünglichen Flächenlabel zuvor ‘gesammelt’ werden.

Betrachten wir zunächst die Kante  $(v_3, v_4)$ . Die beiden Kanten, die zur Bildung von  $(v_3, v_4)$  führen, begrenzen ursprünglich jeweils die Flächen  $f_1$  und  $f_3$  (links). Das führt dazu, daß die beiden Halbkanten von  $(v_3, v_4)$  über ihre *gebundeneFläche* ( $\Delta$ -Zeiger) auf dieselben Flächen verweisen (die Menge  $\{f_1, f_3\}$ ; Mitte). Da wir nur an solchen Kanten interessiert sind, die zwei verschiedene Flächen voneinander abgrenzen, können wir offensichtlich auf  $(v_3, v_4)$  verzichten. Als endgültiges Ergebnis enthält  $\mathcal{D}$  somit nur vier verschiedene Kanten (rechts).

Die Kante  $(v_1, v_3)$  repräsentiert drei ‘Originalkanten’ (links und Mitte) und ihre beiden Halbkanten verweisen auf zwei unterschiedliche Mengen von Flächen, nämlich  $F_1 = \{f_1, f_2, f_1\}$  und  $F_2 = \{f_2, f_1, f_3\}$  (Mitte). Es gibt also zwei Originalflächen,  $f_1$  und  $f_2$ , die auf beiden Seiten von  $(v_1, v_3)$  liegen. Zieht man diese beiden Flächen jeweils von  $F_1$  und  $F_2$  ab, ergeben sich die Mengen der Originalflächen, die von  $(v_1, v_3)$  tatsächlich begrenzt werden und den *gebundeneFläche* ( $\Delta$ -Zeigern) endgültig zugewiesen werden müssen.

Wir müssen die Flächeninformationen also so verwalten, daß nach Aktualisierung der geometrischen Informationen in  $\mathcal{D}$  — d.h. nach Berechnung des durch die Originalkanten der  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$  induzierten planaren Graphen — die Flächenlabel vollständig ‘eingesammelt’ worden sind, so daß sie anschließend korrigiert und überflüssige Kanten entfernt werden können. Wie Abb. 3.31 zeigt, erfordert die Korrektur die Kenntnis aller Flächenlabel, d.h. inklusive von Duplikaten. Flächen, die an eine Halbkante und auch an deren Zwilling gebunden sind, sind an dieser Stelle ‘kollabiert’ (vgl. Kap. 3.5.2.2) und bleiben bei der endgültigen Vergabe der Flächenlabel unberücksichtigt (das bedeutet auch, daß besonders kleine Originalflächen im Ergebnis, d.h. in der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$ , nicht mehr existieren). Den Halbkanten werden schließlich die verbleibenden Flächen als innere Flächen zugewiesen; wenn dann an beide Halbkanten keine Flächen gebunden sind, handelt es sich um eine ‘Sackgasse’, die aus  $\mathcal{D}$  entfernt werden muß. Diese Überlegung ist trivial, so daß wir ohne Beweis den folgenden Hilfssatz angeben:

**Hilfssatz 3.12.** Seien  $h_1$  und  $h_2$  zwei Halbkanten in  $\mathcal{D}$ , so daß  $h_1 = Zwilling(h_2)$ , und seien  $F_1 = innereFläche(h_1)$  und  $F_2 = innereFläche(h_2)$  die Mengen ihrer Flächenlabel nach Abschluß des Scanline-Schwenks. Dann ist  $F_1^! = F_1 \setminus \{F_1 \cup F_2\}$  die Menge der inneren Flächen von  $h_1$  und  $F_2^! = F_2 \setminus \{F_1 \cup F_2\}$  die Menge der inneren Flächen von  $h_2$ .

Die Aktualisierung der Flächenlabel erfordert also zunächst die Zuweisungen  $innereFläche(h_1) := F_1^!$  und  $innereFläche(h_2) := F_2^!$ . Diese Zuweisungen erfolgen für die Halbkanten jeder einzelnen Kante in  $Sub(\mathcal{S}_{1\dots m})$ . Für die Aktualisierung der Flächenlabel reicht es daher aus, den  $Sub(\mathcal{S}_{1\dots m})$

zugrundeliegenden planaren Graphen einmal zu durchlaufen. Mit einem weiteren Durchlauf werden anschließend all jene Kanten aus  $\mathcal{D}$  entfernt, die an keine Flächen mehr gebunden sind. Dabei können gewissermaßen ‘leere’ Eventpunkte in der Eventfolge  $\mathcal{Q}$  entstehen, d.h. Knoten, die an keine Kante mehr gebunden sind. Diese werden anschließend aus  $\mathcal{Q}$  entfernt. Erst dann ist die doppelt-verkettete Kantenliste  $\mathcal{D}$  soweit aktualisiert, daß wir die Flächen tatsächlich erzeugen können.

### 3.5.6 Ein robuster Verschneidealgorithmus

Fassen wir zusammen. Das Herzstück unseres Algorithmus zur simultanen und robusten Verschneidung von  $m$  planaren Subdivisionen ist:

**Algorithmus** SIMULTANEVERSCHNEIDUNG( $\mathcal{S}_{1\dots m}$ )

*Eingabe:* Die doppelt-verketteten Kantenlisten von  $m$  planaren Subdivisionen  $\mathcal{S}_{1\dots m}$ .

*Ausgabe:* Das Ergebnis der Verschneidung von  $\bigcup_{i=1}^m \mathcal{S}_i$  in der doppelt-verketteten Kantenliste  $\mathcal{D}$ .

1. Initialisiere eine leere doppelt-verkettete Kantenliste  $\mathcal{D}$ , eine leere Eventfolge  $\mathcal{Q}$  und eine leere Statusfolge  $\mathcal{T}$ .
2. EVENTFOLGEINITIALISIEREN()
3. ERZEUGEPLANARENGRAPHEN()
4. LÖSCHELEEREKANTEN()
5. LÖSCHELEEREEVENTS()
6. EVENTUPDATE()
7. UMGEBENDEFLÄCHENAKTUALISIEREN()
8. FLÄCHENERZEUGEN()

Verschiedene Operationen, die ursprünglich während eines Scanline-Schwenks durchgeführt worden sind, werden nun nacheinander durchgeführt. Zunächst werden mit EVENTFOLGEINITIALISIEREN die Eventfolge und die doppelt-verkettete Kantenliste  $\mathcal{D}$  initialisiert. Anschließend wird mit ERZEUGEPLANARENGRAPHEN mit Hilfe eines Scanline-Schwenks der planare Graph erzeugt, der durch die Kanten in  $\bigcup_{i=1}^m \mathcal{S}_i$  induziert wird. Dabei werden die Flächeninformationen gesammelt, jedoch nicht weiter bearbeitet. LÖSCHELEEREKANTEN und LÖSCHELEEREEVENTS entfernen alle ‘Sackgassen’ (Kanten, die nicht mindestens zwei verschiedene Flächen begrenzen) und ‘leeren Eventpunkte’ (Knoten, die an weniger als zwei Kanten gebunden sind), die aufgrund einer topologisch konsistenten Verzerrung entstanden sind. Erst danach werden mit EVENTUPDATE die *Vorgänger()*- und *Nachfolger()*-Zeiger der Halbkanten an jedem Knoten aus  $\mathcal{D}$  aktualisiert. In einem zweiten Scanline-Schwenk werden mit UMGEBENDEFLÄCHENAKTUALISIEREN schließlich die Flächeninformationen aktualisiert, so daß für jeden Knoten aus  $\mathcal{D}$  bekannt ist, welche Ursprungsflächen ihn enthalten. Abschließend werden mit FLÄCHENERZEUGEN die Flächen in  $\mathcal{D}$  endgültig erzeugt. Alle diese Operationen sollen im folgenden trotz einiger Wiederholungen in aller Ausführlichkeit dargestellt werden.<sup>37</sup>

Die Prozedur EVENTFOLGEINITIALISIEREN fügt alle Knoten und Kanten aus  $\mathcal{S}_{1\dots m}$  in die Eventfolge  $\mathcal{Q}$  ein. Die Knotenkoordinaten werden dabei durch repräsentierbare Zahlen ersetzt (‘gerundet’).

**Algorithmus** EVENTFOLGEINITIALISIEREN( $\mathcal{S}_{1\dots m}$ )

*Eingabe:* Keine.

*Ausgabe:* Keine.

1. Initialisiere eine jeweils leere doppelt-verkettete Kantenliste  $\mathcal{D}$ , Eventfolge  $\mathcal{Q}$  und Statusfolge  $\mathcal{T}$ .
2. **for each** ( $e \in \bigcup_{i=1}^m \mathcal{S}_i$ ) **do**
3.     Ersetze die Knoten  $v_1$  und  $v_2$  von  $e$  durch die repräsentierbaren Knoten  $v'_1$  und  $v'_2$ .

<sup>37</sup>Die Datenstrukturen  $\mathcal{Q}$ ,  $\mathcal{T}$ ,  $\mathcal{D}$  und  $\mathcal{I}$  sowie der aktuelle Eventpunkt werden im folgenden als bekannt, d.h. als globale Variablen vorausgesetzt.

4.     **if** ( $v'_1 \neq v'_2$ )
- then**
5.             Seien  $h_1$  und  $h_2$  die beiden Halbkanten von  $e$ , und seien  $F_1 = \{\text{innereFläche}(h_1)\}$  und  $F_2 = \{\text{innereFläche}(h_2)\}$ . Setze  $\text{innereFläche}(h_1) := F_1$  und  $\text{innereFläche}(h_2) := F_2$ .
6.             Füge  $v'_1$  und  $v'_2$  in  $\mathcal{Q}$  ein.
7.             KANTEINEVENTFOLGEEINFÜGEN( $e$ )
- end if**
- end for**

Eine Kante muß so in die Eventfolge eingefügt werden, daß die Kante mit beiden Eventpunkten (dem oberen und dem unteren Endpunkt der Kante) gespeichert wird.

**Algorithmus** KANTEINEVENTFOLGEEINFÜGEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* keine.

1. Seien  $v_u$  der obere und  $v_l$  der untere Endpunkt von  $e$ .
2. KANTEMITEVENTPUNKTSPEICHERN( $e, v_u$ )
3. KANTEMITEVENTPUNKTSPEICHERN( $e, v_l$ )

Beim Speichern einer Kante mit einem Eventpunkt muß darauf geachtet werden, daß die Kante an dem Eventpunkt eindeutig von allen anderen Kanten, die an den Eventpunkt gebunden sind, unterscheidbar ist. Die Unterscheidung erfolgt anhand der zyklischen Ordnung dieser Kanten. Ist eine Kante nicht unterscheidbar, liegt eine (Fast-)Überlappung vor, die unmittelbar aufgelöst werden muß. Die Kante, die in die Eventfolge eingefügt wird, wird ggf. auch in die Status- und Intervallfolge eingefügt.

**Algorithmus** KANTEMITEVENTPUNKTSPEICHERN( $e, v$ )

*Eingabe:* Eine Kante  $e$  und ein Eventpunkt  $v$ .

*Ausgabe:* keine.

1. Sei  $\mathcal{C}(v)$  die Folge der Kanten, die in ihrer zyklischen Ordnung um  $v$  mit  $v$  gespeichert sind.
2. Suche  $e$  in  $\mathcal{C}(v)$ . Sei  $e_C$  die Kante, die bei der Suche gefunden wird.
3. **if** ( $e_C$  existiert und  $e_C \neq e$ )
- then**
4.             PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e, e_C$ )
- else**
5.             KANTEINSTATUS EINFÜGEN( $e$ )
6.             Füge  $e$  in  $\mathcal{C}(v)$  und in die doppelt-verkettete Kantenliste  $\mathcal{D}$  ein.
- end if**

Um eine Kante aus der Eventfolge zu löschen, müssen zwei Eventpunkte berücksichtigt werden:

**Algorithmus** KANTEAUSEVENTFOLGELÖSCHEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* keine.

1. Seien  $v_u$  der obere und  $v_l$  der untere Endpunkt von  $e$ .
2. Suche  $v_u$  in der Eventfolge  $\mathcal{Q}$ .
3. Lösche  $e$  aus der Menge der Kanten, die mit  $v_u$  gespeichert sind.
4. Suche  $v_l$  in der Eventfolge  $\mathcal{Q}$ .
5. Lösche  $e$  aus der Menge der Kanten, die mit  $v_l$  gespeichert sind.

Beim Einfügen einer Kante in die Statusfolge  $\mathcal{T}$  muß die Kante immer auch in die Intervallfolge  $\mathcal{I}$  eingefügt werden. Entsprechend erfordert jede Löschoperation in  $\mathcal{T}$  auch eine Löschoperation in  $\mathcal{I}$ :

**Algorithmus** KANTEINSTAUSEINFÜGEN( $e$ )*Eingabe:* Eine Kante  $e$ .*Ausgabe:* keine.

1. **if** ( $e$  ist im aktuellen Eventpunkt in der Statusfolge  $\mathcal{T}$  gültig)  
    **then**
2.     Suche nach  $e$  in der Statusfolge  $\mathcal{T}$ . Sei  $e_{\mathcal{T}}$  die Kante, die bei der Suche gefunden wird.
3.     **if** ( $e_{\mathcal{T}}$  existiert und  $e_{\mathcal{T}} \neq e$ )  
       **then**
4.         PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e, e_{\mathcal{T}}$ ).
- else**
5.         Füge  $e$  in  $\mathcal{T}$  und in  $\mathcal{I}$  ein.
- end if**
- end if**

**Algorithmus** KANTEAUSSTATUSLÖSCHEN( $e$ )*Eingabe:* Eine Kante  $e$ .*Ausgabe:* keine.

1. Lösche  $e$  aus  $\mathcal{T}$  und aus  $\mathcal{I}$ .

Der Algorithmus ERZEUGEPLANARENGRAPHEN ist im wesentlichen eine modifizierte Version der SCHNITTPUNKTSUCHE (vgl. S. 59). Mit ihm wird die Eventfolge durchlaufen und jeder Eventpunkt bearbeitet, wobei auf globale Verletzungen der Invariante geachtet wird: Liegt der neue Schnittpunkt im aktuellen Eventpunkt, wird die Bearbeitung des Eventpunktes wiederholt. Wenn ein Schnittpunkt gefunden wird, der in  $\mathcal{Q}$  vor dem aktuellen Eventpunkt  $v$  liegt, dann muß in  $\mathcal{Q}$  zu diesem Schnittpunkt zurückgeschritten werden. Dabei muß auch die Statusliste  $\mathcal{T}$  entsprechend aktualisiert werden.

**Algorithmus** ERZEUGEPLANARENGRAPHEN( $S$ )*Eingabe:* Keine.*Ausgabe:* Keine.

1. Bestimme den aktuellen Eventpunkt:  $v :=$  der erste Eventpunkt in  $\mathcal{Q}$ .
2. **while** ( $v$  existiert)
3.     Sei der Knoten  $v_s :=$  BEARBEITEEVENTPUNKT( $v$ )
4.     **while** ( $v$  und  $v_s$  existieren und  $v_s \prec_{\mathcal{Q}} v$ )
5.         Sei  $U(v)$  die Menge der Kanten, deren oberer Endpunkt  $v$  ist, und sei  $L(v)$  die Menge der Kanten, deren unterer Endpunkt  $v$  ist.
6.         **for each** ( $e_i \in U(v)$ ) **do**
7.             KANTEAUSSTATUSLÖSCHEN( $e_i$ )
- end for**
8.          $v :=$  der Vorgänger von  $v$  in  $\mathcal{Q}$ .
9.         Weise  $\mathcal{T}$  die neue lokale Ordnung für  $v$  zu
10.         **for each** ( $e_i \in L(v)$ ) **do**
11.             KANTEAUSSTATUSLÖSCHEN( $e_i$ )
- end for**
- end while**
12.     **if** ( $v \prec_{\mathcal{Q}} v_s$ )  
       **then**
13.          $v :=$  der Nachfolger von  $v$  in  $\mathcal{Q}$ .
- end if**
- end while**

Bei der Bearbeitung eines Eventpunktes wird zunächst jeder Eventpunkt den links und rechts benachbarten Kanten zugeordnet. Sich schneidende Kanten werden unmittelbar in ihrem Schnittpunkt

geteilt (dies geschieht in ERZUEGESCHNITTPUNKTEVENT). Anschließend werden jene Kanten, die den aktuellen Eventpunkt enthalten, in ihm geteilt. Die Bestimmung der zu jedem Knoten direkt links benachbarten Kanten (deren Kenntnis ja für die Aktualisierung der Flächeninformationen notwendig ist; vgl. Kap. 3.3.3) erfolgt mit einem separaten Scanline-Schwenk, da sich durch die topologische Verzerrung die Nachbarschaftsverhältnisse ständig ändern können.

**Algorithmus** BEARBEITEEVENTPUNKT( $v$ )

*Eingabe:* Ein Eventpunkt  $v$ .

*Ausgabe:* Ein Schnittpunkt  $v_s$ , wenn  $v_s \prec_Q v$ .

1. Sei  $U(v)$  die Menge der Kanten, deren oberer Endpunkt  $v$  ist, und sei  $L(v)$  die Menge der Kanten, deren unterer Endpunkt  $v$  ist. (Die Kanten in  $U(v) \cup L(v)$  sind mit  $v$  gespeichert.) Sei  $C(v)$  die Menge der Kanten, die  $v$  enthalten und deren Endpunkte ungleich  $v$  sind. (Die Kanten in  $C(v)$  sind in  $\mathcal{T}$  benachbart.)
2. **for each** ( $e_i \in L(v) \cup C(v)$ ) **do**
3.     KANTEAUSSTATUSLÖSCHEN( $e_i$ )
- end for**
4. Weise  $\mathcal{T}$  die neue lokale Ordnung für  $v$  zu
5. Sei  $e$  die in  $\mathcal{T}$  zu  $v$  links benachbarte Kante, und sei  $V_r(e)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  rechts von  $e$  liegen. Wenn  $v$  innerhalb der Bounding-Box von  $e$  liegt, füge  $v$  in  $V_r(e)$  ein.
6. Sei  $e$  die in  $\mathcal{T}$  zu  $v$  rechts benachbarte Kante, und sei  $V_l(e)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  links von  $e$  liegen. Wenn  $v$  innerhalb der Bounding-Box von  $e$  liegt, füge  $v$  in  $V_l(e)$  ein.
7. **for each** ( $e \in U(v) \cup C(v)$ ) **do**
8.     KANTEINSTATUSEIFÜGEN( $e$ )
- end for**
9. **if** ( $U(v) \cup C(v) = \emptyset$ )
- then**
10.     Seien  $e_l$  und  $e_r$  der linke und rechte Nachbar von  $v$  in  $\mathcal{T}$ .
11.     Sei der Knoten  $v_s := \text{ERZUEGESCHNITTPUNKTEVENT}(e_l, e_r)$
12.     **if** ( $v_s$  existiert und  $v_s \preceq_Q v$ )
- then**
- return**  $v_s$
- end if**
- else**
13.     Sei  $e_{ll}$  das am weitesten links stehende Segment aus  $U(v) \cup C(v)$  in  $\mathcal{T}$ .
14.     Sei  $e_l$  der linke Nachbar von  $e_{ll}$  in  $\mathcal{T}$ .
15.     Sei der Knoten  $v_s := \text{ERZUEGESCHNITTPUNKTEVENT}(e_l, e_{ll})$
16.     **if** ( $v_s$  existiert und  $v_s \preceq_Q v$ )
- then**
- return**  $v_s$
- end if**
17.     Sei  $e_{rr}$  das am weitesten rechts stehende Segment aus  $U(v) \cup C(v)$  in  $\mathcal{T}$ .
18.     Sei  $e_r$  der rechte Nachbar von  $e_{rr}$  in  $\mathcal{T}$ .
19.     Sei der Knoten  $v_s := \text{ERZUEGESCHNITTPUNKTEVENT}(e_r, e_{rr})$
20.     **if** ( $v_s$  existiert und  $v_s \preceq_Q v$ )
- then**
- return**  $v_s$
- end if**
- end if**
21. **for each** ( $e \in C(v)$ ) **do**
22.     KANTETEILEN( $e, v$ )
- end for**

Mit der Prozedur KANTETEILEN werden die zu teilende Kante aus  $\mathcal{Q}$  und aus  $\mathcal{D}$  gelöscht und die beiden neu entstandenen Teilkanten in  $\mathcal{Q}$  und  $\mathcal{D}$  eingefügt. Mit den Teilkanten werden sämtliche relevanten Informationen über die benachbarten Knoten und inneren Flächen gespeichert.

**Algorithmus** KANTETEILEN( $e, v$ )

*Eingabe:* Eine Kante  $e$  und ein Knoten  $v$ .

*Ausgabe:* keine.

1. **if** ( $v$  existiert und  $v$  ist kein Endpunkt von  $e$ )  
**then**
2.     KANTEAUSEVENTFOLGELÖSCHEN( $e$ )
3.     KANTEAUSSTATUSLÖSCHEN( $e$ )
4.     Lösche  $e$  aus der doppelt-verketteten Kantenliste  $\mathcal{D}$ .
5.     Seien  $v_u$  der obere und  $v_l$  der untere Endpunkt von  $e$ . Erzeuge zwei neue Kanten  $e_1 = (v_u, v)$  und  $e_2 = (v, v_l)$ .
6.     Setze die *InnereFläche()*-Zeiger der Halbkanten von  $e_1$  und  $e_2$  auf die Mengen, auf die die *InnereFläche()*-Zeiger der Halbkanten von  $e$  verweisen.
7.     Sei  $V_l(e)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  links von  $e$  liegen, sei  $V_r(e)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  rechts von  $e$  liegen, und sei  $V_c(e)$  die Menge aller Knoten, die in  $e$  enthalten sind. (Diese Mengen sind mit  $e$  gespeichert.)
8.     **for each** ( $e_i \in \{e_1, e_2\}$ ) **do**
9.         Sei  $V_l(e_i)$  die Menge aller Knoten aus  $V_l(e)$ , die innerhalb der Bounding-Box von  $e_i$  liegen, sei  $V_r(e_i)$  die Menge aller Knoten aus  $V_r(e)$ , die innerhalb der Bounding-Box von  $e_i$  liegen, und sei  $V_c(e_i)$  die Menge aller Knoten aus  $V_l(e)$ , die innerhalb der Bounding-Box von  $e_i$  liegen.
10.        Speichere  $V_l(e_i)$ ,  $V_r(e_i)$  und  $V_c(e_i)$  mit  $e_i$ .  
**end for**
11.     **for each** ( $e_i \in \{e_1, e_2\}$ ) **do**
12.         Sei  $v_s :=$  FINDEDERIVIERTEENTEILUNGSPUNKT( $e_i$ )
13.         **if** ( $v_s$  existiert)  
**then**
14.             KANTETEILEN( $e_i, v_s$ )  
**else**
15.             KANTEINEVENTFOLGEEINFÜGEN( $e_i$ )
16.             Füge  $e_i$  in  $\mathcal{D}$  ein.  
**end if**
- end for**
- end if**

Die Funktion FINDEDERIVIERTEENTEILUNGSPUNKT gibt genau dann einen Punkt zurück, wenn ein Wechsel einer lokalen Ordnung stattgefunden hat. Ein solcher Punkt entspricht einem vorhandenen (und unverrückbaren) Nagel in unserer Nagelbrett-Analogie (Kap. 3.29).

**Algorithmus** FINDEDERIVIERTEENTEILUNGSPUNKT( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* Einen Knoten  $v$ , wenn  $v$  und  $e$  in einer topologisch inkonsistenten lokalen Ordnung stehen.

1. **for each** ( $v_i \in V_l(e)$ ) **do**
2.     **if** (In  $v_i$  ist  $e \preceq_{\mathcal{T}} v_i$ )  
**then**  
      **return**  $v_i$   
**end if**
- end for**

3. **for each** ( $v_i \in V_T(e)$ ) **do**
4.     **if** (In  $v_i$  ist  $v_i \preceq_T e$ )  
       **then**  
           **return**  $v_i$   
       **end if**
5. **end for**
5. **if** ( $V_c(e) \neq \emptyset$ )  
    **then**  
       **return** Der erste Knoten in  $V_c(e)$ , der nicht ein Endpunkt von  $e$  ist.  
    **end if**

Jede Teilung einer Kante verläuft nun über die Prozedur KANTETEILEN, so z.B. bei der Auflösung von Überlappungen:

**Algorithmus** PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e_1, e_2$ )

*Eingabe:* Zwei sich überlappende Kanten  $e_1$  und  $e_2$ .

*Ausgabe:* keine.

1. Lösche  $e_1$  und  $e_2$  aus  $\mathcal{D}$ .
2. KANTEAUSSTATUSFOLGELÖSCHEN( $e_1$ )
3. KANTEAUSSTATUSFOLGELÖSCHEN( $e_2$ )
4. KANTEAUSEVENTFOLGELÖSCHEN( $e_1$ )
5. KANTEAUSEVENTFOLGELÖSCHEN( $e_2$ )
6. Sei  $t$  der Typ der Überlappung von  $e_1$  und  $e_2$  (vgl. S. 85).
7. **if** ( $t = 3$ )  
    **then**
8.     Seien  $v_i$  und  $v_j$ ,  $v_i \prec_Q v_j$ , die Knoten, in denen  $e_1$  und/oder  $e_2$  geteilt werden müssen, sei  $e_i \in \{e_1, e_2\}$  die Kante, deren oberer Endpunkt nicht  $v_i$  ist, und sei  $e_j \in \{e_1, e_2\}$ ,  $e_i \neq e_j$  die andere Kante. ( $e_i$  ist also die Kante, deren oberer Endpunkt in  $Q$  vor dem von  $e_j$  steht.)
9.     KANTETEILEN( $e_i, v_i$ )
10.    Sei  $e_k$  das Teilstück von  $e_i$ , dessen unterer Endpunkt  $v_i$  ist, und sei  $e_l$  das Teilstück von  $e_i$ , dessen oberer Endpunkt  $v_i$  ist.
11.    KANTEINEVENTFOLGEINFÜGEN( $e_k$ )
12.    PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e_l, e_j$ )
- else if** ( $t = 2$ )  
    **then**
13.    Sei  $v_i$  der gemeinsame Endpunkt von  $e_1$  und  $e_2$ , seien  $e_i$  die längere und  $e_j$  die kürzere der beiden Kanten  $e_1$  und  $e_2$ , und sei  $v_j$  der von  $v_i$  verschiedene Endpunkt von  $e_j$ . ( $v_j$  ist der Punkt, in dem  $e_i$  geteilt werden muß.)
14.    KANTETEILEN( $e_i, v_j$ )
15.    Sei  $e_k$  das Teilstück von  $e_i$ , dessen Endpunkte beide von  $v_i$  verschieden sind, und sei  $e_l$  das Teilstück von  $e_i$ , dessen einer Endpunkt  $v_i$  ist.
16.    KANTEINEVENTFOLGEINFÜGEN( $e_k$ )
17.    PARTIELLEÜBERLAPPUNGAUFLÖSEN( $e_l, e_j$ )
- else if** ( $t = 1$ )  
    **then**
18.    VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN( $e_1, e_2$ )
- end if**

Die Prozedur PARTIELLEÜBERLAPPUNGAUFLÖSEN wurde hier ausführlich dargestellt, um nochmals deutlich zu machen, wie jede Überlappung auf eine vollständige Überlappung reduziert werden kann. Das ist auch insofern von Bedeutung, als daß wir die Vereinigung von Flächeninformationen gewissermaßen mit der Vereinigung zweier deckungsgleicher Kanten 'gleichsetzen' können.

**Algorithmus** VOLLSTÄNDIGEÜBERLAPPUNGAUFLÖSEN( $e_1, e_2$ )

*Eingabe:* Zwei Kanten  $e_1$  und  $e_2$ , die sich vollständig überlappen.

*Ausgabe:* keine.

1. Sei  $v_1$  der obere Endpunkt von  $e_1$  und  $e_2$ , sei  $v_2$  der untere Endpunkt von  $e_1$  und  $e_2$ . Seien  $h_{1d}$  und  $h_{2d}$  die beiden Halbkanten von  $e_1$  und  $e_2$ , deren Ursprung in  $v_1$  liegt, und seien  $h_{1u}$  und  $h_{2u}$  die beiden Halbkanten von  $e_1$  und  $e_2$ , deren Ursprung in  $v_2$  liegt. Seien schließlich  $F_{1d}, F_{2d}, F_{1u}$  und  $F_{2u}$  die Mengen der Flächen, auf die die *innereFläche()*-Zeiger von  $h_{1d}, h_{2d}, h_{1u}$  und  $h_{2u}$  verweisen.
2. Sei  $e = (v_1, v_2)$  eine neue Kante. Sei  $h_d$  die Halbkante von  $e$ , deren Ursprung in  $v_1$  liegt, und sei  $h_u$  die Halbkante von  $e$ , deren Ursprung in  $v_2$  liegt.
3. Setze die *innereFläche()*-Zeiger der Halbkanten von  $e$ , so daß *innereFläche*( $h_d$ ) =  $F_{1d} \cup F_{2d}$  und *innereFläche*( $h_u$ ) =  $F_{1u} \cup F_{2u}$ .
4. Sei  $V_l(e) = V_l(e_1) \cup V_l(e_2)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  und links von  $e_1$  oder links von  $e_2$  liegen, sei  $V_r(e) = V_r(e_1) \cup V_r(e_2)$  die Menge aller Knoten, die innerhalb der Bounding-Box von  $e$  und rechts von  $e_1$  oder rechts von  $e_2$  liegen, und sei  $V_c(e)$  die Menge aller Knoten, die in  $e_1$  oder  $e_2$  enthalten sind. (Diese Mengen sind mit  $e_1$  und  $e_2$  gespeichert.)
5. Speichere  $V_l(e), V_r(e)$  und  $V_c(e)$  mit  $e$ .
6. KANTEINSTAUFÜGEN( $e$ )
7. KANTEINEVENTFOLGEEINFÜGEN( $e$ )
8. Füge die Kante in  $\mathcal{D}$  ein.

Bei der Schnittpunktsuche werden — im Falle eines neu entdeckten Schnittpunktes — die sich schneidenden Kanten unmittelbar geteilt. Der Grund dafür liegt in der Möglichkeit, lokale Verletzungen der Invariante, so wie sie durch die Existenz derivierter Schnittpunkte entstehen, gewissermaßen ‘im Keim zu ersticken’.

**Algorithmus** ERZEUGESCHNITTPUNKTEVENT( $e_1, e_2$ )

*Eingabe:* Zwei Kanten  $e_1$  und  $e_2$ .

*Ausgabe:* Einen neuen Schnittpunkt  $v$ .

1. **if** (Die Endpunkte von  $e_1$  und  $e_2$  sind alle paarweise verschieden)
2. **then**
3. Sei  $v_s$  der Schnittpunkt von  $e_1$  und  $e_2$  nach (Gleichung 3.3).
4. **if** ( $v_s$  existiert)
5. **then**
6. NACHBARKNOTENAKTUALISIEREN( $e_1, e_2, v_s$ )
7. NACHBARKNOTENAKTUALISIEREN( $e_2, e_1, v_s$ )
8. Ersetze die Koordinaten von  $v_s$  durch repräsentierbare Zahlen. (Da in Wirklichkeit nicht  $v_s$  mehrfach kopiert wird, sondern nur der Zeiger auf  $v_s$ , enthalten jetzt auch die Knotenlisten von  $e_1$  und  $e_2$  die Repräsentation von  $v_s$ .)
9. Füge  $v_s$  in  $\mathcal{Q}$  ein.
10. KanteTeilen( $e_1, v_s$ )
11. KanteTeilen( $e_2, v_s$ )
12. DERIVIERTESCHNITTPUNKTEAUFLÖSEN( $e_1, e_2, v_s$ )
13. **return**  $v_s$
14. **end if**
15. **end if**

Die Prüfung auf mögliche Verletzungen der Invariante erfordert die Aktualisierung der benachbarten Knoten von beiden sich schneidenden Kanten  $e_1$  und  $e_2$ , und zwar *vor* einer möglichen ‘Verbiegung’. Mit NACHBARKNOTENAKTUALISIEREN wird zunächst die Menge  $E$  derjenigen Kanten bestimmt, die die Scanlinie schneiden (die also in der Statusfolge  $\mathcal{T}$  und so auch in der Intervallfolge  $\mathcal{I}$  sind) und sowohl unterhalb der Scanlinie als auch oberhalb des Schnittpunktes  $v$  innerhalb der Bounding



Box einer der sich schneidenden Kanten liegen. Nach Hilfssatz 3.8 sind dies die einzigen Kanten, die überprüft werden müssen. Diese Kanten liegen in  $\mathcal{I}$  im Intervall  $[\min(v_1.x, v.x), \max(v_1.x, v.x)]$ , wobei  $v_1$  der obere Endpunkt von  $e_1$  bzw.  $e_2$  ist.<sup>38</sup> Die Kanten  $e_1$  und  $e_2$  gehören nicht zu  $E$ . Der ‘wahre’ Schnittpunkt (der immer unterhalb der Scanlinie liegt) wird in die Knotenlisten aller Kanten in  $E$  eingetragen, und die unteren Endpunkte aller Kanten in  $E$  werden in die Knotenliste einer der sich schneidenden Kanten eingefügt. Im Sinne unserer Nagelbrett-Analogie sind auf diese Weise alle relevanten Nägel bekannt, an denen eine Kante nach ihrer Teilung im repräsentierbaren Schnittpunkt hängenbleiben muß.

**Algorithmus** NACHBARKNOTENAKTUALISIEREN( $e_1, e_2, v$ )

Eingabe: Zwei Kanten  $e_1, e_2$  und einen ‘wahren’ Schnittpunkt  $v$ .

Ausgabe: Keine.

1. Seien  $v_1$  der obere und  $v_2$  der untere Endpunkt von  $e_1$  und sei  $E$  die Menge der Kanten, die in der Intervallfolge im Intervall  $[\min(v_1.x, v.x), \max(v_1.x, v.x)]$  stehen.
2. Lösche  $e_1$  und  $e_2$  aus  $E$ .
3. **for each** ( $e_i \in E$ ) **do**
4.     Seien  $v_u$  der obere Endpunkt und  $v_l$  der untere Endpunkt von  $e_1$ .
5.     **if** ( $v$  liegt innerhalb der Bounding Box von  $e_i$ )  
       **then**
6.         Bestimme die relative Lage von  $v$  zur Kante  $(v_l, v_u)$  nach (Gleichung 3.4).
7.         **if** ( $v$  liegt links von  $(v_l, v_u)$ )  
           **then**
8.             Füge  $v$  in  $V_l(e_i)$  ein.
9.             **else if** ( $v$  liegt rechts von  $(v_l, v_u)$ )  
               **then**
10.                 Füge  $v$  in  $V_r(e_i)$  ein.
11.                 **else**
12.                     Füge  $v$  in  $V_c(e_i)$  ein.
13.             **end if**
14.         **end if**
15.         Sei  $v_{il}$  der untere Endpunkt von  $e_i$ .
16.         **if** ( $v_{il}$  liegt innerhalb der Bounding Box von  $e_1$ )  
           **then**
17.             Bestimme die relative Lage von  $v_{il}$  zur Kante  $(v_2, v_1)$  nach (Gleichung 3.4).
18.             **if** ( $v_{il}$  liegt links von  $(v_2, v_1)$ )  
               **then**
19.                 Füge  $v_{il}$  in  $V_l(e_1)$  ein.
20.                 **else if** ( $v_{il}$  liegt rechts von  $(v_2, v_1)$ )  
                   **then**
21.                     Füge  $v_{il}$  in  $V_r(e_1)$  ein.
22.                     **else**
23.                         Füge  $v_{il}$  in  $V_c(e_1)$  ein.
24.                 **end if**
25.             **end if**
26.         **end if**
27.     **end for**

In ERZUEGESCHNITTPUNKTEVENT werden die Knotenlisten aktualisiert, bevor die Schnittpunktkoordinaten durch repräsentierbare Koordinaten ersetzt werden. Anschließend werden die Kanten im

<sup>38</sup>Strenggenommen können wir das Intervall eingrenzen auf  $[\min(v_t.x, v.x), \max(v_t.x, v.x)]$ , wobei  $v_t$  der Schnittpunkt von  $e_1$  bzw.  $e_2$  mit der Scanlinie ist.

repräsentierbaren Schnittpunkt geteilt und können nun sehr einfach auf lokale Verletzungen der Invariante geprüft werden:

**Algorithmus** DERIVIERTESCHNITTPUNKTEAUFLÖSEN( $e_1, e_2, v$ )

Eingabe: Zwei Kanten  $e_1, e_2$  und einen Knoten  $v$ .

Ausgabe: Keine.

1. Sei  $v_1$  der obere Endpunkt von  $e_1$ , sei  $v_2$  der obere Endpunkt von  $e_2$ , und sei  $E$  die Menge der Kanten, die in der Intervallfolge im Intervall  $[\min(v_1.x, v_2.x, v.x), \max(v_1.x, v_2.x, v.x)]$  stehen.
2. Lösche  $e_1$  und  $e_2$  aus  $E$ .
3. **for each** ( $e_i \in E$ ) **do**
4.     Sei  $v_d := \text{FINDEDERIVIERENTTEILUNGSPUNKT}(e_i)$ .
5.     **if** ( $v_d$  existiert)
6.         **then**
7.             KANTETEILEN( $e_i, v_d$ )
8.         **end if**
9.     **end for**

Zwei sich schneidende Kanten werden damit nach ihrer Teilung wie Schnüre um die Nägel gelegt; tritt dadurch eine neue Verletzung der lokalen Ordnung von Knoten und Kanten auf, wird diese in KANTE-TEILEN rekursiv korrigiert. Durch das Löschen der alten und Wiedereinfügen der korrigierten Kanten in  $\mathcal{Q}$  und ggf.  $\mathcal{T}$  werden auch alle derivierten Überlappungen entdeckt und aufgelöst. Nach jedem neuen Schnittpunkt-Event wird die doppelt-verkettete Kantenliste  $\mathcal{D}$  also so lange ‘zurechtgebogen’, bis keine lokalen Ordnungen mehr verletzt und keine Überlappungen mehr entdeckt werden.

Ist der der doppelt-verketteten Kantenliste  $\mathcal{D}$  zugrundeliegende planare Graph berechnet, müssen die symbolischen Informationen aktualisiert werden. Mit LÖSCHELEEREKANTEN werden die Flächenlabel korrigiert (vgl. Kap. 3.5.5) und sämtliche Kanten, die nicht mindestens zwei verschiedene ‘Originalflächen’ begrenzen, gelöscht.<sup>39</sup>

**Algorithmus** LÖSCHELEEREKANTEN()

Eingabe: Keine.

Ausgabe: Keine.

1. **for each** ( $v \in \mathcal{D}$ ) **do**
2.     Sei  $E$  die Menge der Kanten, die an  $v$  gebunden sind.
3.     **if** ( $E$  enthält mehr als eine Kante)
4.         **then**
5.             **for each** ( $e \in E$ ) **do**
6.                 Seien  $h_1$  und  $h_2$  die beiden Halbkanten von  $e$ . Seien  $F_1 = \text{innereFläche}(h_1)$  und  $F_2 = \text{innereFläche}(h_2)$  die Mengen der Originalflächen, die  $e$  begrenzt.
7.                 Sei  $F = F_1 \cap F_2$ .
8.                 Setze  $F_1 := F_1 \setminus F$  und  $F_2 := F_2 \setminus F$ .
9.                 **if** ( $F_1 \cup F_2 \neq \emptyset$ )
10.                     **then**
11.                         Setze  $\text{innereFläche}(h_1) := F_1$  und  $\text{innereFläche}(h_2) := F_2$ .
12.                     **else**
13.                         KANTELÖSCHEN( $e$ )
14.                     **end if**
15.             **end for**
16.     **end if**
17. **end for**

<sup>39</sup>Man beachte, daß dabei die ungebundenen Flächen mitgezählt werden.

Eine Kante, die keine Originalfläche begrenzt, wird also unmittelbar gelöscht. Dabei wird ggf. auch ein ganzer Kantenzug gelöscht, wenn nämlich durch die Entfernung der Kante eine ‘Sackgasse’ entstanden ist (d.h. wenn ein Knoten existiert, an den genau eine Kante gebunden ist):

**Algorithmus** KANTELÖSCHEN( $e$ )

*Eingabe:* Eine Kante  $e$ .

*Ausgabe:* Keine.

1. Seien  $v_1$  und  $v_2$  die beiden an  $e$  gebundenen Knoten.
2. KANTEAUSEVENTFOLGELÖSCHEN( $e$ )
3. Lösche  $e$  aus  $\mathcal{D}$ .
4. Sei  $E_1$  die Menge aller an  $v_1$  gebundenen Kanten.
5. **if** ( $E_1$  enthält genau eine Kante  $e_i$ )  
**then**
6.     KANTELÖSCHEN( $e_i$ )
7.     **end if**
7. Sei  $E_2$  die Menge aller an  $v_2$  gebundenen Kanten.
8. **if** ( $E_2$  enthält genau eine Kante  $e_i$ )  
**then**
9.     KANTELÖSCHEN( $e_i$ )
10.    **end if**

Durch das Löschen von Kanten können leere Eventpunkte entstehen (d.h. Knoten, an die keinen Kanten gebunden sind), die in einem weiteren Durchlauf aus der Eventfolge entfernt werden:

**Algorithmus** LÖSCHELEEREEVENTS()

*Eingabe:* Keine.

*Ausgabe:* Keine.

1. **for each** ( $v \in \mathcal{D}$ ) **do**
2.     Sei  $E$  die Menge der Kanten, die an  $v$  gebunden sind.
3.     **if** ( $E = \emptyset$ )  
**then**
4.         Lösche  $v$  aus  $\mathcal{Q}$  und aus  $\mathcal{D}$
5.     **end if**
6. **end for**

Die doppelt-verkettete Kantenliste  $\mathcal{D}$  enthält jetzt alle Knoten und (Halb-)Kanten, die zu dem der planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$  zugrundeliegenden Graphen gehören. Die Halbkantenzüge, die die Flächen in  $Sub(\mathcal{S}_{1\dots m})$  begrenzenden, können dann ‘zusammengesetzt’ werden:

**Algorithmus** EVENTUPDATE()

*Eingabe:* Keine.

*Ausgabe:* Keine.

1. **for each** ( $v \in \mathcal{D}$ ) **do**
2.     Sei  $E$  die Menge der Kanten, die an  $v$  gebunden sind.
3.     Aktualisiere die *Vorgänger()*- und *Nachfolger()*-Zeiger der Halbkanten aller Kanten in  $e$  nach dem ‘Linksabbieger-Prinzip’ wie in Kap. 3.2.3.1 (S. 67) besprochen.
4. **end for**

Um für jeden Knoten die Originalflächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  zu bestimmen, die ihn enthalten, muß für jeden Knoten die links benachbarte Kante bekannt sein (vgl. Satz 3.4 in Kap. 3.3.3). Da diese Information

aufgrund der ständigen ‘Verbiegungen’ und ‘derivierten Überlappungen’ nicht schon in ERZEUGE-PLANARENGRAPHEN aktualisiert werden konnte, ist jetzt ein zweiter Scanline-Schwenk notwendig.

**Algorithmus** UMGEBENDEFLÄCHENAKTUALISIEREN( $S$ )

Eingabe: Keine.

Ausgabe: Keine.

1. Bestimme den aktuellen Eventpunkt:  $v :=$  der erste Eventpunkt in  $\mathcal{Q}$ .
2. **while** ( $v$  existiert)
3. Sei  $U(v)$  die Menge der Kanten, deren oberer Endpunkt  $v$  ist, und sei  $L(v)$  die Menge der Kanten, deren unterer Endpunkt  $v$  ist. (Die Kanten in  $U(v) \cup L(v)$  sind mit  $v$  gespeichert.)
4. **for each** ( $e_i \in L(v)$ ) **do**
5. KANTEAUSSTATUSLÖSCHEN( $e_i$ )
6. **end for**
7. Weise  $\mathcal{T}$  die neue lokale Ordnung für  $v$  zu
8. Sei  $e_l$  die direkt links zu  $v$  benachbarte Kante. ( $e_l$  ist in  $\mathcal{T}$  der Voränger von  $v$ .) Sei  $v_u$  der obere Endpunkt von  $e_l$ .
9. **for each** ( $e \in U(v)$ ) **do**
10. KANTEINSTAUSEINFÜGEN( $e$ )
11. **end for**
12. Sei  $E(v_u)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , die  $v_u$  enthalten (sie sind mit  $v_u$  gespeichert, da  $v_u \prec_{\mathcal{Q}} v$ ).
13. Sei  $L(v_u)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , deren äußere Grenzen durch  $v_u$  verlaufen und gegen den Uhrzeigersinn orientiert sind.
14. Sei  $B(v)$  die Menge der Flächen in  $\bigcup_{i=1}^m \mathcal{S}_i$ , deren Grenzen durch  $v$  verlaufen.
15. Sei  $E(v) = \{E(v_u) \cup L(v_u)\} \setminus B(v)$ .
16. Speichere  $E(v)$  mit  $v$ .
17.  $v :=$  der Nachfolger von  $v$  in  $\mathcal{Q}$ .
18. **end while**

Jetzt liegen alle Informationen vor, so daß die Flächen in  $Sub(\mathcal{S}_{1\dots m})$  bestimmt und die Originalflächen aus  $\bigcup_{i=1}^m \mathcal{S}_i$  als Label zugewiesen werden können.

**Algorithmus** FLÄCHENERZUEGEN()

Eingabe: Keine.

Ausgabe: Keine.

1. Erzeuge einen leeren Flächeneintrag in  $\mathcal{D}$  für die ungebundene Fläche
2. **for each** (Eventpunkt  $v$  in  $\mathcal{Q}$ ) **do**
3. Sei  $U(v)$  die Menge der Halbkanten  $h$ , so daß  $Ursprung(h) = v$  und  $v \prec_{\mathcal{Q}} Ziel(h)$
4. **for each** (Halbkante  $h \in U(v)$ ,  $innereFläche(h) \notin \mathcal{D}$ ) **do**
5. Sei  $H$  der Halbkantenzug, der in  $h$  beginnt.
6. **if** ( $H$  ist gegen den Uhrzeigersinn orientiert) **then**
7. Erzeuge eine neue Fläche  $f$  mit  $äußereBegrenzung(f) := h$  und füge  $f$  in  $\mathcal{D}$  ein.
8. Umlaufe  $H$  und bestimme dabei die Vereinigungsmenge  $F$  aller Flächen, auf die die  $innereFläche()$ -Zeiger aller Halbkanten  $h_i \in H$  verweisen und innerhalb derer die Ursprünge aller Halbkanten  $h_i \in H$  liegen:
$$F = \bigcup_{h_i \in H} \left( innereFläche(h_i) \cup umgebendeFlächen(Ursprung(h_i)) \right)$$
9. Speichere  $F$  mit  $f$ .
10. Setze für jede Halbkante  $h_i \in H$   $innereFläche(h_i) := f$  und aktualisiere  $h_i$  in  $\mathcal{D}$ .

```

else
11.     Sei  $h'$  eine Halbkante.
12.     Sei  $v_l$  der am weitesten links liegende Knoten in  $H$ .
13.     Sei  $h_d :=$  die Halbkante, die direkt links von  $v_l$  liegt und für die  $Ursprung(h_d) \prec_Q Ziel(h_d)$ 
        ist.
14.      $H :=$  der Kantenzug, der in  $h_d$  beginnt.
15.      $h' := h$ .
16.     while ( $H$  existiert und  $H$  ist im Uhrzeigersinn orientiert)
17.          $h' := h_d$ .
18.          $v_l :=$  der am weitesten links liegende Knoten in  $H$ .
19.          $h_d :=$  die Halbkante, die direkt links von  $v_l$  liegt und für die  $Ursprung(h_d) \prec_Q Ziel(h_d)$ 
            ist.
20.          $H :=$  der Kantenzug, der in  $h_d$  beginnt.
    end while
21.     Sei  $f := innereFläche(h')$ .
22.     Sei  $H$  der Kantenzug, der in  $h$  beginnt.
23.     Füge  $h$  in  $innereBegrenzung(f)$  ein und aktualisiere den Eintrag von  $f$  in  $\mathcal{D}$ .
24.     Setze für jede Halbkante  $h_i \in H$   $innereFläche(h_i) := f$  und aktualisiere  $h_i$  in  $\mathcal{D}$ .
end if
end for
end for

```

### 3.5.7 Zeitaufwand

Weche Zeit benötigt nun der Algorithmus für die Berechnung der durch  $m$  planare Subdivisionen  $\mathcal{S}_{1\dots m}$  induzierten  $\varepsilon$ -nahen planaren Subdivision  $Sub(\mathcal{S}_{1\dots m})$ ? Über die Wahl der Gitterweite, d.h. des minimalen Abstandes zwischen zwei in  $Sub(\mathcal{S}_{1\dots m})$  repräsentierbaren Knoten, haben wir die Möglichkeit, die Komplexität  $k$  von  $Sub(\mathcal{S}_{1\dots m})$  nach Belieben zu steuern. Der Zeitaufwand für die Initialisierung der Eventfolge  $\mathcal{Q}$  und der doppelt-verketteten Kantenliste  $\mathcal{D}$  liegt – wegen der Projektion aller Knoten aus  $\bigcup_{i=1}^m \mathcal{S}_i$  auf repräsentierbare Knoten in EVENTFOLGEINITIALISIEREN – jetzt in  $O(n \log k + l \log k)$ , wobei  $n$  die Summe der Komplexitäten von  $\mathcal{S}_{1\dots m}$  und  $l$  die Anzahl der sich überlappenden Kanten ist. Man beachte, daß bei hinreichend großer Gitterweite  $k < n$  ist.

Läßt man den Aufwand zur Aktualisierung der ‘Nagelbrettgeometrie’ unberücksichtigt, dann liegt der Zeitaufwand für die Bestimmung des  $\bigcup_{i=1}^m \mathcal{S}_i$  zugrundeliegenden planaren Graphen mit ERZUEGEPLANARENGRAPHEN in  $O(k \log k + l \log k)$  (vgl. Satz 3.6). Der zusätzliche Aufwand besteht darin, die Intervallfolge  $\mathcal{I}$  zu führen und bei jedem neuen Schnittpunkt auf Verletzungen der Invariante zu prüfen. Einfüge- und Löschooperationen in  $\mathcal{I}$  benötigen  $O(\log k)$  Zeit, wobei  $k$  die Komplexität von  $Sub(\mathcal{S}_{1\dots m})$  ist, was genau dem Aufwand entspricht, der für die Aktualisierung der Statusfolge  $\mathcal{T}$  notwendig ist. Da Kanten in  $\mathcal{I}$  und  $\mathcal{T}$  immer gleichzeitig eingefügt bzw. gelöscht werden, verdoppelt sich die benötigte Zeit und liegt damit für die Aktualisierung von  $\mathcal{I}$  und  $\mathcal{T}$  insgesamt in  $O(k \log k)$ .

Für jede Prüfung auf eine lokale Verletzung der Invariante müssen in  $\mathcal{I}$  drei Intervallabfragen durchgeführt werden, und zwar in NACHBARKNOTENAKTUALISIEREN und DERIVIERTESCHNITTPUNKTEAUFLÖSEN. Jede Suche erfordert  $O(\log k + r)$  Zeit, wobei  $r$  die Größe der Antwort (die Anzahl der gefundenen Kanten) ist (vgl. Ottmann, Widmayer 1993, S. 503). Die Überprüfung der gefundenen Knoten und Kanten in DERIVIERTESCHNITTPUNKTEAUFLÖSEN erfordert zusätzlich  $O(r)$  Zeit. Werden nun in DERIVIERTESCHNITTPUNKTEAUFLÖSEN lokale Verletzungen der Invariante gefunden, wird eine Kante entsprechend häufig geteilt. Die dafür erforderliche Zeit liegt in  $O(d \log k)$ , wobei  $d$  die Anzahl der derivierten Schnittpunkte ist. Die Gesamtlaufzeit erhöht sich also durch Über-

prüfung und Auflösung derivierter Schnittpunkte um  $O(Ik \log k)$ , wobei  $I$  die Anzahl der Schnittpunkte in  $\bigcup_{i=1}^m \mathcal{S}_i$  ist.

Beim Auftreten einer globalen Verletzung der Invariante muß in der Eventfolge zurückgeschritten werden. Neben  $\mathcal{Q}$  müssen sowohl  $\mathcal{T}$  als auch  $\mathcal{I}$  entsprechend aktualisiert werden, so daß jeder Rückschritt  $O(\log k)$  Zeit erfordert. Die insgesamt zusätzlich benötigte Zeit für das ‘Zurück-Ć und wieder ‘Vorwärtsschreiten’ liegt ebenfalls in  $O(Ik \log k)$ .

Das Löschen leerer Kanten mit LÖSCHELEEREKANTEN erfordert für jede Kante in  $Sub(\mathcal{S}_{1\dots m})$  die Bestimmung der Schnittmengen von Flächenlabeln und liegt daher in  $O(k \log k)$ . LÖSCHELEEREVENTS benötigt  $O(k)$ , EVENTUPDATE wiederum  $O(k \log k)$  Zeit. Der zur Aktualisierung der umgebenden Flächen mit UMGEBENDEFLÄCHENAKTUALISIEREN erforderliche Zeitaufwand liegt in  $O(k \log k)$  und die Erzeugung der Flächeneinträge schließlich in  $O(k)$ .

**Satz 3.13.** *Seien  $\mathcal{S}_{1\dots m}$   $m$  planare Subdivisionen der Komplexitäten  $n_1, n_2, \dots, n_m$ , und sei  $n := \sum_{i=1}^m n_i$ . Die zur robusten Berechnung der  $\varepsilon$ -nahen planaren Subdivision von  $\bigcup_{i=1}^m \mathcal{S}_i$  benötigte Zeit liegt in  $O(n \log k + Ik \log k + l \log k)$ , wobei  $k$  die Komplexität der Verschneidung,  $l$  die Anzahl der sich überlappenden Kanten und  $I$  die Anzahl der Schnittpunkte in  $\bigcup_{i=1}^m \mathcal{S}_i$  ist.*

# Kapitel 4

## Anwendungsbeispiele

Abschließend soll die Leistungsfähigkeit des vorgeschlagenen Verfahrens anhand zweier Beispiele demonstriert werden. In Kap. 4.1 wird zunächst eine erdraumbezogene (geographische), in Kap. 4.2 anschließend eine merkmalsräumliche Anwendung vorgestellt. Beide Beispiele dienen ausschließlich zur Demonstration der Verfahrensweise und sollen dem Leser ermöglichen, nachzuvollziehen, wie mit Hilfe planarer Subdivisionen sowohl der Erdoberfläche als auch jeder anderen Euklidischen Ebene rationiert werden kann. Ziel ist nicht die weitergehende inhaltliche Auseinandersetzung mit den angerissenen Themen, weshalb die Beispielanwendungen auch nicht mehr als einen einführenden Charakter haben. Vielmehr sollen die Möglichkeiten aufgezeigt werden, die Flächenverschneidungen und damit letztendlich auch GIS bieten können.

### 4.1 Erreichbarkeit sozialer Infrastruktur

Eines der Leitziele sowohl der Raumordnungspolitik als auch der Stadtplanung ist die Gleichwertigkeit der Lebensbedingungen in allen Teilräumen (Haubner 1982). Der Bremer Senator für Bau, Verkehr und Stadtentwicklung nennt entsprechend als eines seiner Leitziele, die „ausgewogene Versorgung der Bevölkerung mit sozialer, kultureller und Bildungsinfrastruktur“ zu sichern (Freie Hansestadt Bremen, Der Senator für Bau, Verkehr und Stadtentwicklung 1998, S. 14). Aus stadtplanerischer Sicht ist natürlich die Überprüfung der Wirksamkeit des Planungsinstrumentariums von entscheidender Bedeutung. Am Beispiel der Bremer Neustadt soll hier gezeigt werden, wie eine Evaluation planerischer Leitbilder mit Hilfe einer multiplen Flächenverschneidung durchgeführt werden kann. Dabei soll es an dieser Stelle natürlich nur darum gehen, einen ersten Einstieg in die Thematik zu finden und gewissermaßen eine Raumabstraktion der Bremer Neustadt zu entwickeln, die es erlaubt, als Diskussionsvorlage die weitere thematische Auseinandersetzung anzustoßen.

Abb. 4.1 zeigt die Standorte von Bildungs- und sozialen Einrichtungen in der Bremer Neustadt. Dazu zählen

- Einrichtungen für Kinder (Kindertagesstätten, Kindergärten etc.),
- Einrichtungen für Jugendliche (Jugendheime und -clubs etc.),
- Einrichtungen für alte Menschen (Altentagesstätten, Altersheime etc.),
- kirchliche Einrichtungen (Seelsorge, Gebetskreise etc.),
- Bildungseinrichtungen (Grund-, Haupt- und Realschulen, Gymnasien etc.),
- sonstige Einrichtungen (betreutes Wohnen, Sozialämter etc.).

Diese Einrichtungen sollen im folgenden kurz als soziale Infrastruktur bzw. Einrichtungen bezeichnet werden. Die Versorgung mit sozialen Einrichtungen kann in einer ersten Näherung als Erreichbarkeit von Versorgungsstandorten verstanden werden. Gut erreichbar ist ein Standort dann, wenn er vom Wohnstandort in wenigen Minuten zu Fuß erreicht werden kann. Für unser Beispiel nehmen wir an, daß dies einer Luftlinienentfernung von maximal 500 m entspricht. Die Wohnstandorte liegen in Form von Baublöcken vor, für die jeweils die Flächengröße und die Einwohnerzahl für das Jahr 1996 bekannt sind. Der Versorgungsgrad mit sozialer Infrastruktur kann dann auf zweierlei Weise definiert werden, erstens als Durchmischung und zweitens als Konzentration sozialer Einrichtungen, d.h. erstens als Anzahl der maximal sechs verschiedenen Standorttypen und zweitens als Anzahl der Standorte insgesamt, die vom Wohnstandort aus gut erreichbar sind. Letzteres ist vor allem dann interessant, wenn die Bevölkerungsdichte mit berücksichtigt wird, da dann die Dichte der sozialen Einrichtungen mit der Dichte der Wohnbevölkerung verglichen werden kann (vgl. Abb. 4.2).

Die geometrische Formulierung des Problems lautet also wie folgt: Gegeben sind die Baublockflächen sowie die Punktkoordinaten der Standorte sozialer Einrichtungen. Der Bereich, von dem aus ein solcher Standort gut erreichbar ist, läßt sich durch eine Kreisfläche mit einem Radius von 500 m um den Standort herum angeben (vgl. Abb. 4.3; die Kreisflächen sind durch Zwölfecke angenähert). Zu bestimmen ist die Menge aller Teilflächen, die durch die Begrenzungen aller Baublöcke und aller Kreise induziert werden. Für jede Teilfläche muß anschließend bekannt sein, welche Kreise und Baublöcke ursprünglich übereinanderliegen, so daß Rechenoperationen, wie die Bestimmung von Bevölkerungs- oder Standortdichten, im Rückgriff auf die Ausgangsdaten möglich sind. Mit anderen Worten: Ziel ist die Berechnung einer planaren Subdivision der Bremer Neustadt, mit der die Neustadt in einzelne, sich nicht überschneidende *Zellen* partitioniert wird, für die jeweils die Informationen über den Bevölkerungsbesatz und über die von dort erreichbaren sozialen Einrichtungen bekannt sind. Dafür werden reale Flächen (die Baublöcke) mit fiktiven Flächen (den Einzugsbereichen) verschnitten.

Wie bei jeder erdräumlichen Beschreibung der Verteilung sozialer bzw. infrastruktureller Phänomene können wir es nicht vermeiden, einige modellhafte Annahmen zu treffen. Zum einen wird nämlich unterstellt, die Nachfrage nach sozialen Einrichtungen sei unabhängig von der Lage des Wohnstandortes über die gesamte Neustadt konstant. Zum anderen müssen wir annehmen, daß die Bevölkerung innerhalb eines Baublocks gleichmäßig verteilt ist. Aufgrund der relativ hohen erdräumlichen Auflösung der Baublöcke stellt die zweite Annahme kaum ein Problem dar. Die erste Annahme ist dagegen durchaus mit Skepsis zu betrachten, denn sie unterstellt letztendlich die Isotropie der Neustädter Sozialstruktur. Sie läßt sich nur vermeiden, wenn wir zusätzlich z.B. Ergebnisse einer sozialräumlichen Analyse mit in unsere Untersuchung einbeziehen. Das ist prinzipiell problemlos möglich, muß aber an dieser Stelle aufgrund der mangelhaften Datenlage unterbleiben.

#### 4.1.1 Durchmischung und Konzentration sozialer Infrastruktur

Abb. 4.4 zeigt die Durchmischung der sozialen Infrastruktur in der Bremer Neustadt. Das dunkelgraue Areal im Zentrum entspricht dem Gebiet, von dem aus alle sechs verschiedenen Typen von sozialen Einrichtungen gut erreicht werden können; von jeder beliebigen Stelle innerhalb dieses Areals sind also sechs verschiedene Einrichtungen nicht weiter als 500 m Luftlinie entfernt. Mit zunehmender Distanz zum Zentrum nimmt die Durchmischung ab, im nordwestlichen Stadtteil (Woltmershausen) liegt allerdings ein weiterer Mischungsschwerpunkt. Die Durchmischung der sozialen Infrastruktur entspricht damit in etwa der Verteilung der Bevölkerung (Abb. 4.2).

Ein deutlich anderes Bild zeigt die Konzentration der Standorte sozialer Einrichtungen (Abb. 4.5). Die Bereiche, von denen aus die meisten Einzelstandorte gut erreichbar sind, liegen südwestlich vom Mischungsschwerpunkt. Der Grund dafür ist in erster Linie eine Häufung von Einrichtungen südlich



des Neustädter Zentrums, die auf der Standortkarte (Abb. 4.1) nicht zu erkennen ist, da viele Einrichtungen im selben Gebäude untergebracht sind. Wesentlich deutlicher ist auch der Unterschied zwischen dem Neustädter Zentrum und Woltmershausen im Nordwesten ausgeprägt. Ein Vergleich der Standortkonzentration (Abb. 4.5) mit der Bevölkerungsdichte (Abb. 4.2) macht allerdings deutlich, daß die Versorgungsdichte der sozialen Infrastruktur großräumig im wesentlichen der Bevölkerungsverteilung entspricht.

#### 4.1.2 Relativer Versorgungsgrad

Es bleibt die Frage, ob die unterschiedliche Erreichbarkeit sozialer Einrichtungen der Verteilung der Bevölkerung entspricht. Um dies zu klären, berechnen wir den relativen Versorgungsgrad, und zwar auf Basis der Erreichbarkeit der Einzelstandorte (vgl. Abb. 4.5). Dabei gehen wir wie folgt vor: Die Anzahl der Einrichtungen, die von den verschiedenen Teilflächen der Bremer Neustadt gut erreicht werden können, liegt zwischen 0 und 28. Daraus läßt sich die relative Erreichbarkeit als Prozentwert ermitteln, so daß 28 genau 100% entspricht. Auf die gleiche Weise bestimmen wir die relative Bevölkerungsdichte, so daß die maximale Bevölkerungsdichte einer Zelle 100% entspricht. Eine ausgewogene Versorgung liegt dann vor, wenn die relative Erreichbarkeit in etwa gleich der relativen Bevölkerungsdichte ist. Abb. 4.6 zeigt das Ergebnis. Die Zellen wurden dabei so klassifiziert, daß eine Abweichung von  $\pm 10\%$ -Punkten als durchschnittlich gilt (weiße Bereiche). Liegt die relative Erreichbarkeit um mehr als 10%-Punkte über der relativen Bevölkerungsdichte, sprechen wir von einer relativen Überversorgung (hellgraue Areale), im umgekehrten Fall von einer relativen Unterversorgung (dunkelgraue Areale).

Wenn auch, wie oben angemerkt, die Konzentration sozialer Einrichtungen großräumig in etwa der Bevölkerungsverteilung entspricht, können wir nun feststellen, daß es kleinräumig deutliche Differenzen gibt. Im Vergleich zur gesamten Bremer Neustadt sind in unserem Sinne die Randlagen der Hauptwohngebiete unterversorgt, aber auch weite Bereiche von Woltmershausen und sogar ganz zentral gelegene Zellen. Erstaunlich ist zudem, daß weite Bereiche mit sehr hohen Bevölkerungsdichten tatsächlich überversorgt sind. Hier liegen offensichtlich Konzentrationseffekte vor, deren Erklärung allerdings einer weitergehenden Untersuchung vorbehalten ist. Der relative Versorgungsgrad läßt sich schließlich auch nutzen, um potentielle Standorte neuer sozialer Einrichtungen zu bestimmen.

Man beachte, daß die vorgestellten Ergebniskarten (Abb. 4.4- 4.6) auf ein und derselben Verschneidung basieren. Dies ist möglich, da die Verschneidung im Vektorformat vorliegt und über die Kenntnis der Originalflächen auf die Ausgangsdaten Bezug genommen werden kann (vgl. Kap. 1.1 und Kap. 3). Insbesondere in Abb. 4.6 wird deutlich, welche Möglichkeiten diese Form der Partitionierung der Ebene bietet. Stehen z.B. Daten über die Altersstruktur auf Baublockniveau zur Verfügung, können wesentlich spezifischere Problemlagen visualisiert werden, z.B. die Versorgung älterer Bewohner mit Altentagesstätten oder die Versorgung von Jugendlichen mit entsprechenden Freizeiteinrichtungen u.v.m. Mit diesem Hinweis soll jedoch nicht einer gedankenlosen Visualisierung Vorschub geleistet werden. Vielmehr wird nur angeregt, die vorgeschlagene Technik als eine mögliche Form der Visualisierung zu begreifen, um so auf graphischem Wege relativ detaillierte, übersichtliche und v.a. leicht verständliche Diskussionsvorlagen erarbeiten zu können.



**Abbildung 4.1** Standorte von sozialen und Bildungseinrichtungen in der Bremer Neustadt. Als Kartengrundlage ist die Struktur der Baublöcke dargestellt.



**Abbildung 4.2** Die Bevölkerungsdichte in der Bremer Neustadt. Die Bevölkerungsdichte ist auf Basis der Baublockflächen dargestellt (Einwohner je ha).



**Abbildung 4.3** Die Flächen der Baublöcke und Einzugsbereiche in der Bremer Neustadt. Das geometrische Problem bei der Evaluation einer ausgewogenen Versorgung der Bevölkerung mit sozialer Infrastruktur besteht darin, die planare Subdivision zu bestimmen, die durch die Flächen der Baublöcke und der Einzugsbereiche sozialer Einrichtungen induziert wird.



**Abbildung 4.4** Durchmischungsgrad von sozialen und Bildungseinrichtungen in der Bremer Neustadt. Als Durchmischungsgrad wird die Anzahl der Standorte unterschiedlicher Standorttypen bezeichnet, die von einer Zelle gut erreichbar sind.



**Abbildung 4.5** Konzentration von sozialen und Bildungseinrichtungen in der Bremer Neustadt. Als Konzentration wird die Anzahl der Standorte insgesamt bezeichnet, die von einer Zelle aus gut erreichbar sind.



**Abbildung 4.6** Relativer Versorgungsgrad mit sozialen und Bildungseinrichtungen in der Bremer Neustadt. Der relative Versorgungsgrad bezeichnet die Differenz *relative Bevölkerungsdichte* - *relative Standortkonzentration*.

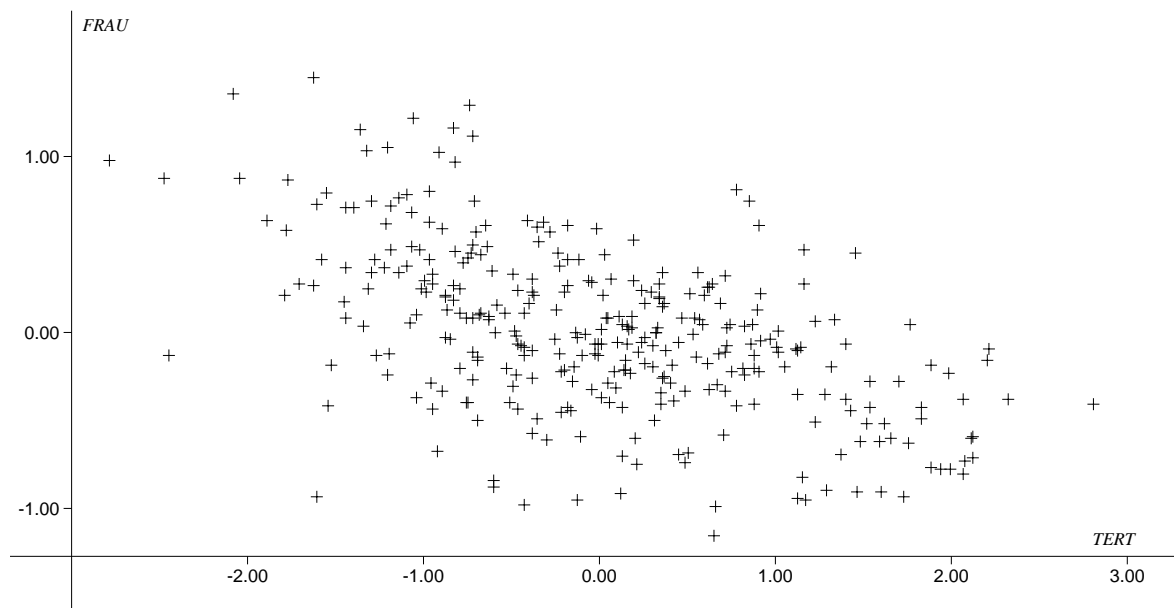


Abbildung 4.7 Das Streudiagramm der Variablen *TERT* und *FRAU*.

## 4.2 Visuelle Regressionsanalyse

In Kontrast zur geographischen Anwendung, bei der es ja auch schon um die simultane Verschneidung zahlreicher z.T. fiktiver Flächen ging, soll jetzt eine merkmalsräumliche Anwendung des Verschneideverfahrens vorgestellt werden. Dabei geht es insbesondere darum, die didaktische Funktion von Flächenverschneidungen herauszustellen. Als Beispiel dient eine zweifache Regressionsanalyse, mit der die Frage beantwortet werden soll, inwieweit die Arbeitslosigkeit von Frauen in Westdeutschland von der lokalen Arbeitsplatzstruktur abhängig ist. Die zu prüfende These ist, daß der Anteil der Frauen an den Arbeitslosen mit zunehmendem Tertiärisierungsgrad abnimmt. Die Daten wurden der Laufenden Raumbesichtigung entnommen (Bundesforschungsanstalt für Landeskunde und Raumordnung 1995). Auf Basis der westdeutschen Kreise und kreisfreien Städte stehen uns folgende Variablen zur Verfügung:

- *TERT* = Anteil der sozialversicherungspflichtig Beschäftigten im Tertiären Sektor an allen sozialversicherungspflichtig Beschäftigten am 30.6.1993 (in %) und
- *FRAU* = Anteil der arbeitslosen Frauen an den Arbeitslosen am 30.9.1993 (in %).

Beide Variablen wurden standardisiert. Wie man im Streudiagramm (Abb. 4.7) leicht sieht, besteht der vermutete negative Zusammenhang tatsächlich. Als Ergebnis einer formalen Regression von *FRAU* nach *TERT* erhält man entsprechend die Regressionsgleichung

$$FRAU = -0.54 \cdot TERT$$

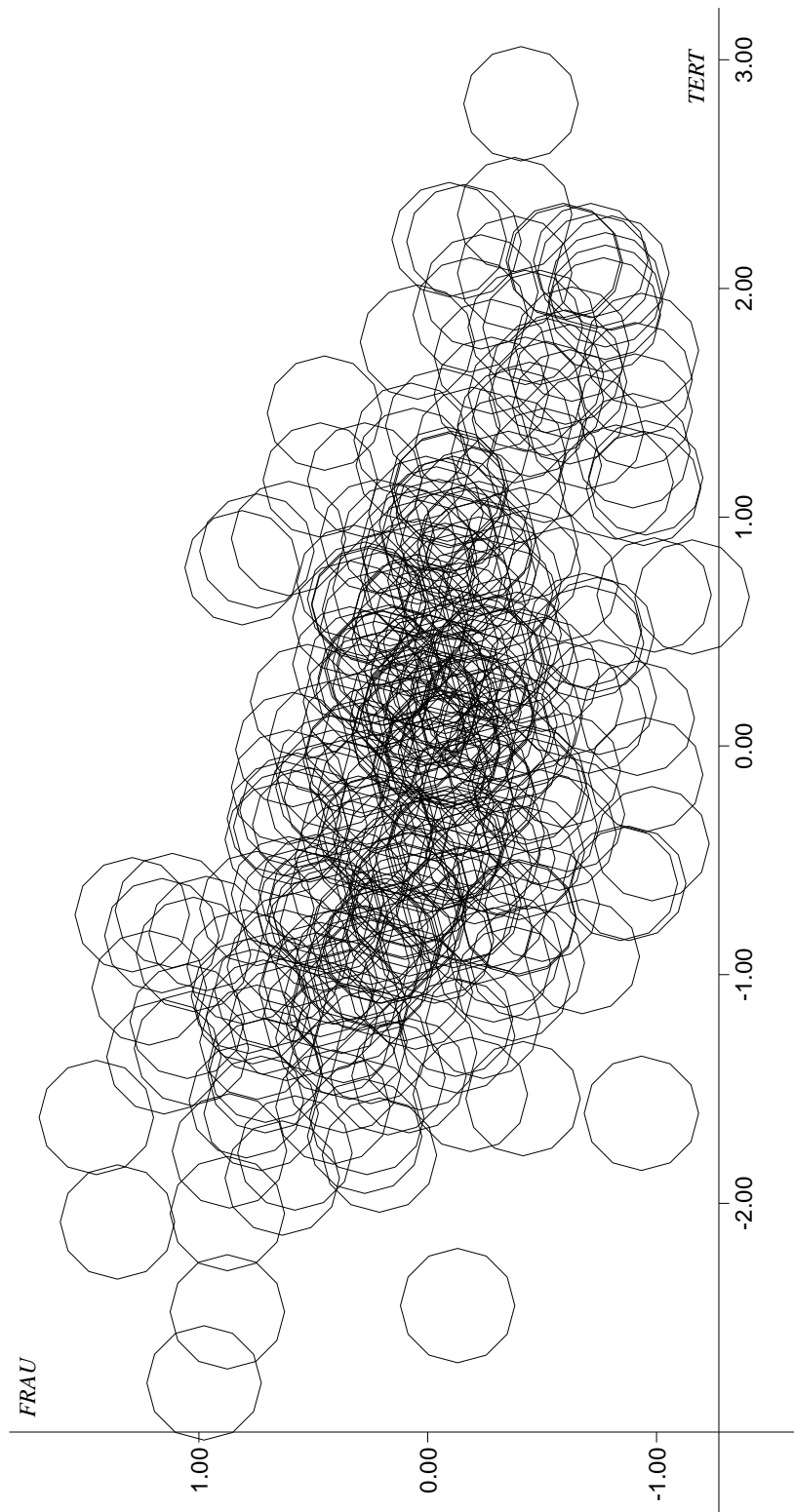
Ein großes Problem bei der Vermittlung des Verfahrens zur Schätzung der Regressionsgleichung nach dem Gausschen Prinzip der kleinsten Quadrate stellt die Minimierung der Abstandsquadrate dar. In der Regel ist es hilfreich, vor dem Nachvollziehen dieses Schritts eine bildhafte Vorstellung davon zu



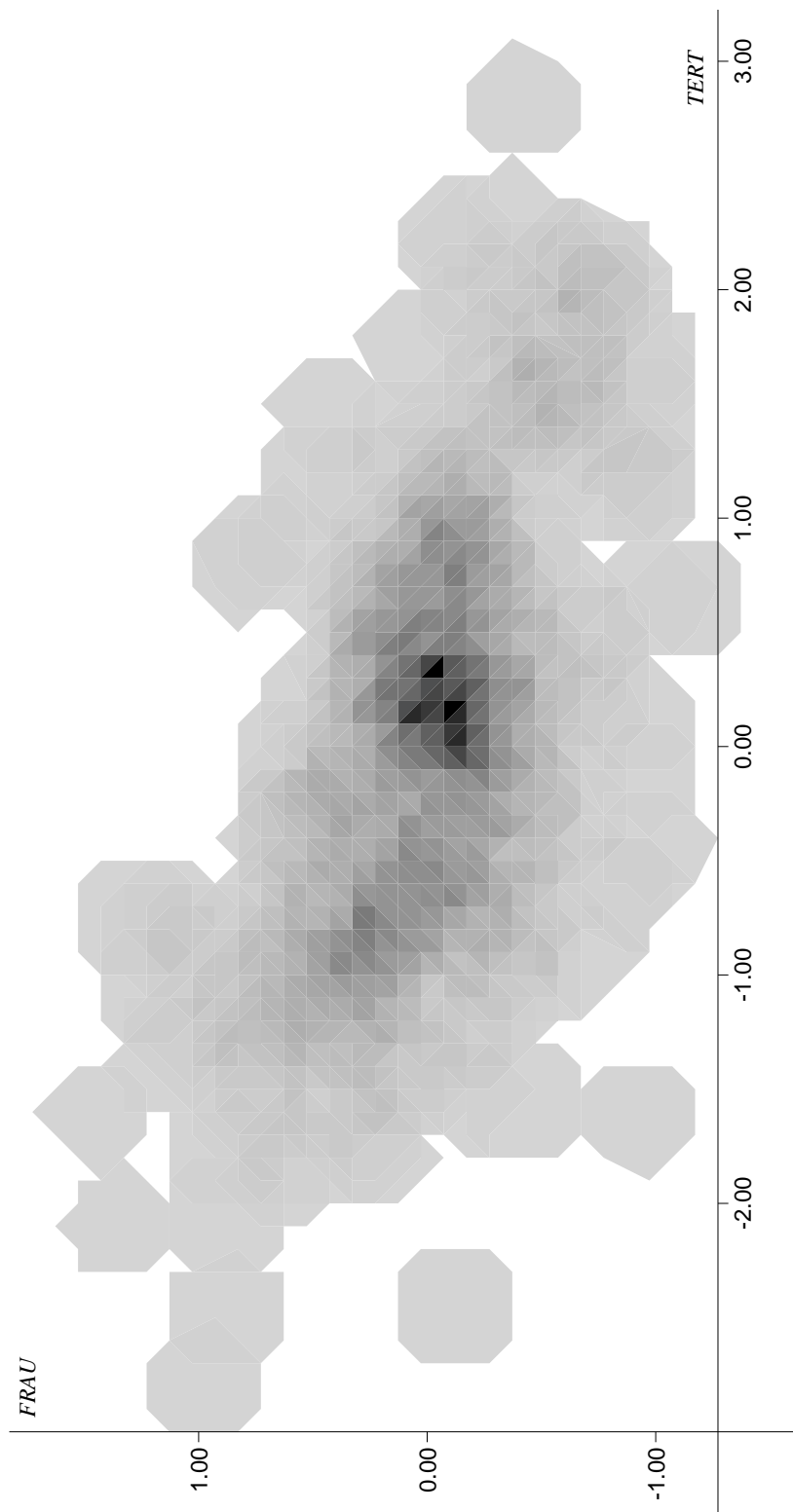
entwickeln. Das gelingt leicht, indem das Verfahren anhand eines erdräumlichen Beispiels erläutert wird, z.B. dem Bau einer (geraden) ICE-Hochgeschwindigkeitstrasse unter der Bedingung, daß alle Kühe, die auf den Weiden innerhalb des Bauabschnitts stehen, insgesamt möglichst nah an der Trasse grasen sollen (wodurch der Ausblick der Fahrgäste optimiert werden soll; die Tatsache, daß man gegen eine solche Vorgabe schwere inhaltliche Bedenken anmelden kann, tut der Anschaulichkeit nach den Erfahrungen des Verfassers keinen Abbruch – eher das Gegenteil ist der Fall). Ungeachtet der fragwürdigen Vereräumlichung kann auf diese Weise das abstrakte Problem der Optimierung einer Geraden in das Problem einer Trassenoptimierung umgewandelt werden, und das ist in der Regel wesentlich leichter nachzuvollziehen (vgl. Kap. 2). Wenn nämlich den Kühen ein gewisser Aktionsradius unterstellt wird, besteht das Problem der Trassenplanung darin, die Trasse so zu legen, daß möglichst viele dieser Aktionsräume geschnitten werden.

Wem diese Metaphorik zu weit geht, kann natürlich am Streudiagramm festhalten und die Idee der Aktionsradien in Form von Flächen übernehmen, in der ein Datenpunkt der Grundgesamtheit mit einer gewissen Wahrscheinlichkeit liegt. Aus den Datenpunkten werden dann kreisförmige Datenflächen (Abb. 4.8; im vorgestellten Beispiel wurde ein Durchmesser von 0,5 Standardabweichungen gewählt) und das Problem der Geradenoptimierung läßt sich analog zum obigen Beispiel beschreiben. Die Anzahl der sich überschneidenden Datenflächen kann nun natürlich mit Hilfe des allgemeinen Flächenverschneideverfahrens bestimmt werden, und zwar als die planare Subdivision, die durch die Grenzen aller Datenflächen induziert wird. Abb. 4.9 zeigt das Ergebnis einer solchen merkmalsräumlichen Verschneidung, die Lage der Regressionsgeraden zeichnet sich als dunkler Bereich deutlich ab.

Um diese Gleichung nach der Methode der kleinsten Quadrate zu schätzen oder den standardisierten Regressionskoeffizienten als Korrelationskoeffizienten zu interpretieren, müssen mehrere Voraussetzungen erfüllt sein, insbesondere, was die Normalität der beiden Variablen betrifft (vgl. Bahrenberg, Giese, Nipper 1985, S. 154 ff.; dies 1992, S. 34). Dem Streudiagramm können wir die Häufigkeitsverteilung der beiden Variablen nicht direkt entnehmen, und wir müssen strenggenommen zunächst einen Test auf Binormalverteilung durchführen – ein Aufwand, der in der Regel unterbleibt. Auch diese Häufigkeitsverteilung kann in Abb. 4.9 unmittelbar abgelesen werden. Neben der didaktischen Komponente bietet die merkmalsräumliche Flächenverschneidung also auch die Möglichkeit, formale Voraussetzungen visuell auf einfachem Wege zu überprüfen.



**Abbildung 4.8** Das Streudiagramm der Variablen *TERT* und *FRAU* mit kreisförmigen Datenflächen. Die Abbildung ist aus Platzgründen um 90° gedreht.



**Abbildung 4.9** Eine visuelle Regressionsanalyse. Dargestellt ist die planare Subdivision, die durch die Kreisflächen in Abb. 4.8 induziert wird. Mit der Anzahl der sich überschneidenden Kreisflächen nimmt die Intensität der Grautöne zu, die maximale Anzahl sich überschneidender Flächen ist 35 (schwarz). Als Gitterweite wurde 0,1 gewählt (vgl. Kap. 3.5). Der Verlauf der Regressionsgeraden ist als dunkler Bereich unmittelbar zu erkennen. Die bivariate Häufigkeitsverteilung kann anhand der Gradienten der Grauwerte abgelesen werden. Die Abbildung ist aus Platzgründen um  $90^\circ$  gedreht.

# Kapitel 5

## Zusammenfassung

Der unbestreitbare Erfolg Geographischer Informationssysteme wird häufig damit begründet, daß die Probleme und Fragestellungen, die mit GIS bearbeitet werden, relevant sind für die Erhaltung der physischen und sozialen Lebensgrundlagen des Menschen. Diese Einschätzung trifft für zahlreiche GIS-Anwendungen sicherlich zu und wurde in der vorliegenden Arbeit in keiner Weise in Frage gestellt. Was dagegen hinterfragt wurde, ist die Forschungslogik derer, die theoretische oder wissenschaftliche Grundlagen für GIS propagieren. Dabei zeigte sich, daß das, was als wissenschaftliche *Theoriebildung* thematisiert wird, letztendlich immer nur die Beforschung der GIS-*Technik* zum Inhalt hat.

In Kap. 1 wurde gezeigt, daß dieser Widerspruch ganz in der Tradition der Geographie als wissenschaftliche Fachdisziplin steht. Neben der Anwendungsrelevanz besteht eine wichtige Funktion von GIS darin, die Einheit der Geographie zu gewährleisten. Die angestrebte wissenschaftstheoretische Grundlegung von GIS scheint nämlich zu gelingen, indem GIS als Möglichkeit verstanden wird, (endlich!) die traditionelle geographisch-ganzheitliche Welt zu beforschen. Da dieser Anspruch heute, angesichts einer hochgradig ausdifferenzierten Wissenschaftslandschaft, von einer einzigen Disziplin nicht einzulösen ist, werden fachfremde Disziplinen in die GIS-Arbeit mit eingebunden, deren Integration über die 'Räumlichkeit' des mit GIS zu untersuchenden Objektes gelingt. Jedoch bedeutet 'Räumlichkeit' für jede Disziplin etwas anderes: Während es Geographen um die Verteilung etc. von Phänomenen usf. auf der Erdoberfläche geht, steht für Kognitionspsychologen u.a. die Bedeutung von Bildern bei der mentalen Verarbeitung von Informationen im Zentrum des Interesses, Ethnologen geht es dagegen um den Ausdruck sozialer Hierarchien u.v.m. Was bleibt, ist nicht die Entwicklung einer einheitlichen Raumtheorie, sondern die Hinwendung zu einer gemeinsamen Technik. Die Weiterentwicklung der GIS-Technik – und darin liegt die Bedeutung dieser eigenartigen argumentativen Schleife für die integrative Funktion von GIS – wird so scheinbar mit theoretischem Gehalt angereichert, der jedoch einer näheren Überprüfung kaum standhält.

Wer sich wissenschaftlich mit GIS beschäftigen will, muß einen anderen Zugang zum Raumbe-griff finden. Dabei soll nicht bestritten werden, daß 'raumbezogene' Abbildungen, so wie sie mit GIS erzeugt werden, sehr wohl sinnvoll sein können. In der vorliegenden Arbeit wurde allerdings die Ansicht vertreten, daß die Nutzungseffizienz solcher Abbildungen nicht primär darin besteht, erdräumlich verortbare Phänomene abzubilden, sondern darin, daß beliebige Phänomene verräumlicht werden, um so die Kommunikation zwischen unterschiedlich informierten Beteiligten zu erleichtern. Die Funktionsweise solcher einfachen Raumabstraktionen im Sinne von geometrischen Abstraktionen wurde von Klüter (1986) aus sozialwissenschaftlicher Perspektive geklärt.

Der Schwerpunkt in Kap. 2 wurde dagegen auf die Aufarbeitung kognitionspsychologischer Erkenntnisse gelegt. Die Idee dabei war, gewissermaßen hinter die sozialwissenschaftlichen Kulissen zu schauen und der Frage nachzugehen, wie und warum geometrische Abstraktionen es den Menschen in

der Regel erleichtern zu verstehen, zu kommunizieren und zu deduzieren. Es zeigte sich, daß die Funktion geometrischer Abstraktionen offensichtlich darin gesehen werden muß, die mentale Modellierung abstrakter Sachverhalte zu erleichtern. Als Ursache dafür werden sowohl evolutionär vererbte als auch tradierte Fähigkeiten angesehen, die für das Überleben der Menschen als auf der Erdoberfläche existierende Individuen notwendig sind. Hierzu gehören Orientierungsfähigkeiten sowie Techniken, die eine erdräumliche Orientierung erleichtern, z.B. das Zeichnen von Wegbeschreibungen und Landkarten. Es gibt zahlreiche Hinweise darauf, daß die Modelle, die wir beim abstrakten Denken aufbauen, bildhaft sind und damit auch geometrische sowie topologische Eigenschaften aufweisen, die, obwohl sie im Einzelfall verzerrt sein können, den Verhältnissen auf der Erdoberfläche entsprechen. In diesem Sinne könnte die abstrakte Ableitung eines logischen Widerspruchs der mentalen Konstruktion eines unmöglichen Körpers entsprechen.

Letztendlich bedeutet das, daß wir zumindest in vielen Bereichen 'räumlich' denken. Der Raum, den wir uns dabei vorstellen, entspricht dem erdoberflächlichen Raum, in dem wir uns bewegen und uns körperlich wahrnehmen, und er ist mit Dingen angefüllt, von denen nicht zwei an ein und derselben Erdstelle existieren können. Auf der Erdoberfläche ist das Gesetz vom ausgeschlossenen Dritten körperlich erfahrbar und als ein Nebeneinander von Teilen der Erdoberfläche bzw. der auf ihr verteilten Objekte gewissermaßen materialisiert: Der Bremer Roland steht entweder vor dem Bremer Rathaus oder er steht nicht dort, eine dritte Möglichkeit existiert nicht. Die Entwicklung einer zweiwertigen Logik und entsprechender Deduktionsregeln folgt damit erdoberflächlichen Verhältnissen. Umgekehrt gilt: Je abstrakter die Verhältnisse sind, über die deduziert wird, desto notwendiger ist ihre Transformation in räumliche Strukturen, die zu denen der Erdoberfläche isomorph sind. Jede geometrische Abstraktion ist daher genau dann sinnvoll, wenn sie die logischen Eigenschaften dessen, was abstrahiert wird, bewahrt, und gleichzeitig die topologischen Eigenschaften erdoberflächlicher erfahrbarer Verhältnisse widerspiegelt.

Aus diesem Grunde ist das Venn-Diagramm das Idealbild einer geometrischen Abstraktion, denn es entspricht genau unserer Erfahrung der Erdoberfläche als Menge nebeneinanderliegender und sich nicht überlappender Teilflächen. Die Erfahrung der Erdoberfläche als planare Subdivision gilt für uns natürlich immer nur bei der Betrachtung von Phänomenen genau einer Kategorie, z.B. Bodennutzung oder Bodentyp, ein Rasonieren auf Grundlage beider Kategorien erfordert dagegen zunächst die Erzeugung der Planarität. Genau diese Operation bietet GIS in Form der Flächenverschneidung, die als geographische Variante des Venn-Diagramms gedeutet werden kann. Auch die umgekehrte Deutung, daß nämlich das Venn-Diagramm Ausdruck unserer Konzeptionalisierung der Erdoberfläche ist, auf der wir uns bewegen, ist möglich und nach Ansicht des Verfassers auch wahrscheinlicher.

Die Form der Visualisierung, die mit der Verschneidung von Flächen ermöglicht wird, bietet daher einen ungemein anschaulichen Zugang zu abstrakteren Formen der Datenanalyse. In diesem Zusammenhang stellt sich jedoch das Problem, daß die bekannten Flächenverschneidungsalgorithmen nicht die simultane Verschneidung von beliebig vielen planaren Subdivisionen erlauben. Die notwendige Verallgemeinerung der bekannten Verfahren wurde ausführlich in Kap. 3 besprochen. Neben der Frage, wie die geometrischen, topologischen und attributiven Informationen bei einer simultanen Verschneidung von mehr als zwei planaren Subdivisionen aktualisiert werden müssen, wurde auch das Problem der Robustheit des Verfahrens gelöst.

Wer mit GIS arbeitet, ist in der Regel an der Produktion von Schnitt-, Vereinigungs- oder Differenzflächen als Teilmengen der Erdoberfläche interessiert. Flächenverschneidungen mögen in vielen Fällen sinnvoll sein; in dem Moment, in dem Teile der Erdoberfläche bzw. die Art und Weise ihrer Überlagerung als *Ursache* für physische und insbesondere soziale Phänomene verstanden werden, ist jedoch höchste Vorsicht geboten (vgl. Kap. 1). Das geographische Beispiel, das in Kap. 4.1 gegeben wurde, hat einen gänzlich anderen Charakter, denn erdräumliche Größen werden nicht als Erklärung,

sondern einzig als eine mögliche *Form* der Beschreibung genutzt. Die Stärke dieser Form liegt insbesondere darin, daß der visuelle Eindruck gleichermaßen als Beschreibung stadtplanerischer Problemlagen und als simple Lösung eines Standortallokationsproblems gedeutet werden kann. Sie hat damit vor allem auch didaktischen Wert, denn das Lösungsverfahren ist – ganz im Gegensatz zur Lösung eines *p*-Median-Problems – trivial. Kap. 4.2 zeigte schließlich, daß neben erdräumlichen noch andere sinnvolle geometrische Beschreibungen existieren, die in ähnlicher Weise ein raumbezogenes Rasonieren unterstützen.

In der vorliegenden Arbeit ging es zu einem großen Teil auch um technische Fragen. Im Unterschied zu gängigen Akademisierungsstrategien fußten die hier erörterten technischen Fragen jedoch auf einer Auseinandersetzung darüber, *was* und *warum was* mit GIS sinnvoll zu bearbeiten ist. Die herausgestellte nicht-geographische Funktionalität von GIS in Form eines ‘Geometrischen Imaginations-Systems’, d.h. eines Computersystems zur Erzeugung von Diagrammen, die einer erdräumlichen Logik und Metaphorik entsprechen, ist nach Ansicht des Verfassers eine sinnvolle Möglichkeit, GIS zum Gegenstand der Forschung zu machen, und kann damit tatsächlich ein fruchtbares Betätigungsfeld darstellen für alle, die *über* GIS forschen wollen.

# Literaturverzeichnis

- Aangeenbrug, R.T. (1991): A critique of GIS. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991): Geographical Information Systems. Principles and Applications. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical. 101-107.
- Abler, R.; Adams, J.S.; Gould, P. (1971): Spatial Organization. The Geographer's View of the World. Englewood Cliffs, NJ.: Prentice-Hall.
- Acredolo, L.P. (1978): Development of spatial orientation in infancy. *Developmental Psychology* 14, 224-234.
- Albrecht, J. (1996): Universal GIS-Operations. A Task-Oriented Systematization of Data Structure-Independent GIS Functionality Leading Towards a Geographic Modeling Language. Vechta: Institut für Strukturfor- schung und Planung in agrarischen Intensivgebieten. Hochschule Vechta. = Mitteilungen 23.
- Allen, J.F. (1983): Maintaining knowledge about temporal intervals. *Communications of the ACM* 26 (11), 832-843.
- Anderson, M. (Hrsg.) (1997): Reasoning with diagrammatic representations. AAAI Fall Symposium on Reasoning with Diagrammatic Representations, MIT, Cambridge, Mass., 8.11.-10.11.1997. Menlo Park, CA: AAAI Press. = AAAI Technical Report 97-03.
- Anooshian, L.J. (1996): Diversity within spatial cognition: Strategies underlying spatial knowledge. *Environment and Behavior* 28 (4), 471-493.
- Anselin, L.; Getis, A. (1993): Spatial statistical analysis and geographic information systems. In: Fischer, M.M.; Nijkamp, P. (Hrsg.) (1993) a.a.O. 35-49.
- Antonio, F. (1994): Faster line segment intersection. In: Kirk, D. (Hrsg.) (1994): Graphics Gems III. Boston: Academic Press. 199-202.
- Ardila, A. (1993): Historical evolution of spatial abilities. *Behavioral Neurology* 6 (2), 83-87.
- Arnheim, R. (1969): Visual Thinking. Berkeley, CA: University of California Press.
- Arvo, J. (Hrsg.) (1994): Graphics Gems II. Boston: Academic Press.
- Atkinson, R.C.; Shiffrin, R.M. (1968): Human memory: A proposed system and its control process. In: Spence, K.W. (Hrsg.) (1968): The Psychology of Learning and Motivation: Advances in Research and Theory. Vol 2. New York: Academic Press. 89-195.
- Aumann, G.; Spitzmüller, K. (1993): Computerorientierte Geometrie. Mannheim, Leipzig, Wien, Zürich: BI-Wissenschaftsverlag. = Reihe Informatik 89.
- Baddeley, A.D. (1968): A three-minute reasoning test based on grammatical transformations. *Psychonomic Science* 10, 341-342.
- Baddeley, A.D. (1976): The Psychology of Memory. New York: Basic Books.
- Baddeley, A.D. (1986): Working Memory. Oxford: Clarendon Press.
- Baddeley, A.D. (1997): Human Memory: Theory and Practice. Hove: Psychology Press.
- Baddeley, A.D.; Hitch, G. (1974): Working memory. In: Bower, G.A. (Hrsg.) (1974): Recent Advances in Learning and Motivation. Vol. 8. New York: Academic Press.
- Bahrenberg, G. (1987): Über die Unmöglichkeit von Geographie als „Raumwissenschaft“ – Gemeinsamkeiten in der Konstituierung von Geographie bei A. Hettner und D. Bartels. In: Bahrenberg, G.; Deiters, J.; Fischer, M.M.; Gaebel, W.; Hard, G.; Löffler, G. (Hrsg.) (1987): Geographie des Menschen. Dietrich Bartels zum Gedenken. Bremen: Universität Bremen. = Bremer Beiträge zur Geographie 11. 225-239.

- Bahrenberg, G.; Deiters, J.; Fischer, M.M.; Gaebe, W.; Hard, G.; Löffler, G. (Hrsg.) (1987): *Geographie des Menschen. Dietrich Bartels zum Gedenken*. Bremen: Universität Bremen. = Bremer Beiträge zur Geographie 11.
- Bahrenberg, G.; Fischer, M.M. (Hrsg.) (1984): *Theorie und Quantitative Methodik in der Geographie*. Bremen: Universität Bremen. = Bremer Beiträge zur Geographie und Raumplanung 5.
- Bahrenberg, G.; Fischer, M.M. (Hrsg.) (1986): *Theoretical and Quantitative Geography. Proceedings of the Third European Colloquium Held at Augsburg, 13.9.-17.9.1982*. Bremen: Universität Bremen. = Bremer Beiträge zur Geographie und Raumplanung 8.
- Bahrenberg, G.; Giese, E.; Nipper, J. (1985): *Statistische Methoden in der Geographie*. Band 1. Stuttgart: Teubner. = Teubner Studienbücher Geographie.
- Bahrenberg, G.; Giese, E.; Nipper, J. (1992): *Statistische Methoden in der Geographie*. Band 2. Stuttgart: Teubner. = Teubner Studienbücher Geographie.
- Barker, S.M. (1995): Towards a topology for computational geometry. *Computer Aided Design* 27 (4), 311-318.
- Barlow, H. (1990): What does the brain see? How does it understand? In: Barlow, H.; Blakemore, C.; Weston-Smith, M. (Hrsg.) (1990): *Images and Understanding. Thoughts about Images. Ideas about Understanding*. Cambridge, New York: Cambridge University Press. 5-25.
- Barlow, H.; Blakemore, C.; Weston-Smith, M. (Hrsg.) (1990): *Images and Understanding. Thoughts about Images. Ideas about Understanding*. Cambridge, New York: Cambridge University Press.
- Bartelme, N. (1995): *Geoinformatik. Modelle, Strukturen, Funktionen*. Berlin: Springer.
- Bartels, D. (1968): *Zur wissenschaftstheoretischen Grundlegung einer Geographie des Menschen*. Wiesbaden: Franz Steiner. = Erdkundliches Wissen 19.
- Bartels, D. (1970a): Einleitung. In: Bartels, D. (Hrsg.) (1970b): a.a.O. 13-45.
- Bartels, D. (Hrsg.) (1970b): *Wirtschafts- und Sozialgeographie*. Köln, Berlin: Kiepenheuer & Witsch. = Neue Wissenschaftliche Bibliothek 35.
- Bartels, D. (1974): Schwierigkeiten mit dem Raumbegriff. *Geographica Helvetica* 29 (2/3, Beiheft).
- Batty, M. (1997): The Computable City. *International Planning Studies* 2 (2), 155-173.
- Bauer, M.I.; Johnson-Laird, P.N. (1993): How diagrams can improve reasoning. *Psychological Science* 4, 372-378.
- Berry, B.J.L. (1967): *Geography of market centers and retail distribution*. Englewood Cliffs, N.J.: Prentice Hall. = Foundation of Economic Geography Series.
- Berry, B.J.L.; Marble, D.F. (Hrsg.) (1968): *Spatial Analysis. A Reader in Statistical Geography*. Englewood Cliffs: Prentice Hall.
- Bill, R.; Fritsch, D. (1991): *Grundlagen der Geo-Informationssysteme*. Band 1: Hardware, Software, Daten. Karlsruhe.
- Bittner, T.; Frank, A.U. (1997): An Introduction to the application of formal theories to GIS. In: Dollinger, F.; Strobl, J. (Hrsg.) (1997)a.a.O. 11-22.
- Bobek, H. (1957): Gedanken über das logische System der Geographie. *Mitteilungen der Geographischen Gesellschaft Wien* 99, 122-145.
- Bobek, H.; Schmithüsen, J. (1949): Die Landschaft im logischen System der Geographie. *Erdkunde* 3 (2/3), 112-120.
- Boole, G. (1854): *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*. London: MacMillan.
- Bower, G.A. (Hrsg.) (1974): *Recent Advances in Learning and Motivation*. Vol. 8. New York: Academic Press.
- Braine, M.D.S. (1978): On the relation between the natural logic of reasoning and standard logic. *Psychological Review* 85, 1-21.
- Brassard, G.; Bratley, P. (1993): *Algorithmik. Theorie und Praxis*. Attenkirchen: Wolfram.
- Bromley, D.F.; Selman, J.F. (1993): Geographic information usage in a local authority: The human environment for GIS. *Area* 25 (3), 228-236.



- Broschart, J. (1995): The social perception of space. Non-spatial determinants of the use of directionals in Tongan (Polynesia). In: Frank, A.U.; Kuhn, W. (Hrsg.) (1995): *Spatial Information Theory: A Theoretical Basis for GIS. Proceedings of the International Conference COSIT '95, Semmering, Austria, 21.9.-23.9.1995*. Berlin: Springer. = *Lecture Notes in Computer Science* 988. 443-462.
- Brückler, M. (1996): Bodenpreisstudien für ÖROK und Land Salzburg. *SIR-Info* (1996), 1, 6-7.
- Brückler, M.; Dumfarth, E. (1996): Erstellung von Bodenpreiskarten durch GIS-gestützte räumliche Interpolation. Methode und weiterführende Analysemöglichkeiten. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 39-50.
- Bryant, D.J.; Lanca, M.; Tversky, B. (1995): Spatial concepts and perception of physical and diagrammed scenes. *Perceptual and Motor Skills* 81, 531-546.
- Bugelski, B.R.; Kidd, E.; Segmen, J. (1968): Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology* 76, 69-73.
- Bundesforschungsanstalt für Landeskunde und Raumordnung (BFLR) (Hrsg.) (1995): *Laufende Raumbeobachtung. Aktuelle Daten zur Entwicklung der Städte, Kreise und Gemeinden 1992/93*. Bonn: BFLR. = *Materialien zur Raumentwicklung* 67.
- Burrough, P.A. (1986): *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford: Clarendon Press. = *Monographs on Soil and Resources Survey* 12.
- Buttenfield, B.; Ganter, J.H. (1990): Visualization and GIS: What should we see? What might we miss? In: Brassel, K.; Kishimoto, H. (Hrsg.) (1990): *Proceedings of the Fourth International Symposium on Spatial Data Handling, Zürich (Vol. 1)*. Zürich. 307-316.
- Buttimer, A. (1982): Musing on the Helicon: Root metaphors and Geography. *Geografiska Annaler* 64 B (2), 89-96.
- Byrne, R.M.J.; Johnson-Laird, P.N. (1989): Spatial reasoning. *Journal of Memory & Language* 28, 564-575.
- Byrne, R.W. (1979): Memory for urban geography. *Quarterly Journal of Experimental Psychology* 31, 147-154.
- Cameron, S.; Yap, C.K. (1992): Refinement methods for geometric bounds in constructive solid geometry. *ACM Transactions on Graphics* 11 (1), 12-39.
- Campari, I.; Frank, A.U. (Hrsg.) (1995): *Experiential realism and its applications to geographic space*. Wien: Technische Universität Wien. = *GeoInfo* 1.
- Campbell, J. (1994): *Past, Space, and Self*. Cambridge, MA.: MIT Press.
- Car, A.; Frank, A.U. (1994): Hierarchisches räumliches Schließen - Allgemeine Prinzipien. In: Dollinger, F.; Strobl, J. (Hrsg.) (1994) a.a.O. 151-162.
- Carlson-Radvansky, L.A.; Irwin, D.E. (1993): Frames of reference in vision and language: Where is above? *Cognition* 46, 223-244.
- Carol, H. (1956): Zur Diskussion um Landschaft und Geographie. *Geographica Helvetica* 11 (2), 111-132.
- Chen, Z. (1995): Analogical transfer: From schematic pictures to problem solving. *Memory & Cognition* 23 (2), 255-269.
- Christaller, W. (1933): *Die zentralen Orte in Süddeutschland. Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Jena (wiederaufgelegt Darmstadt 1968).
- Coad, P. (1992): Object-oriented patterns. *Communications of the ACM* 35 (9), 152-159.
- Colle, H.A.; Welsh, A. (1976): Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior* 15, 17-32.
- Conrad, R. (1964): Acoustic confusion in immediate memory. *British Journal of Psychology* 55, 429-432.
- Conrad, R.; Hull, A.J. (1964): Information, acoustic confusion and memory span. *British Journal of Psychology* 55, 429-432.
- Conway, M.A. (Hrsg.) (1997): *Cognitive Models of Memory*. Hove: Psychology Press. = *Studies in Cognition*.
- Cooper, L.A. (1995): Varieties of visual representation: How are we to analyse the concept of mental image? *Neuropsychologica* 33 (11), 1575-1582.

- Cooper, L.A.; Podgorny, P. (1976): Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception & Performance* 2, 503-514.
- Cosmides, L.; Tooby, J. (1996): Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition* 58, 1-73.
- Couclelis, H. (1991): Requirements for planning relevant GIS: a spatial perspective. *Papers in Regional Science* 70, 9-19.
- Craik, K. (1943): *The Nature of Explanation*. Cambridge: Cambridge University Press.
- Cressie, N.; Helterbrand, J.D. (1994): Multivariate spatial statistical models. *Geographical Systems* 1 (3), 179-188.
- Cummins, R. (1996): *Representations, Targets, and Attitudes*. Cambridge, MA.: MIT Press.
- Curiel, J.M.; Radvansky, G.A. (1998): Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24 (1), 202-214.
- Czernanka, M. (1996): Spatial Decision Support Systems in Naturschutz und Landschaftspflege? Umsetzungsaspekte für die raumbezogene Planung. *Laufener Seminarbeiträge* (1996), 4, 21-27.
- Dangermond, J. (1990): A classification of software components commonly used in geographic information systems. In: Peuquet, D.J.; Marble, D.F. (Hrsg.) (1990): *Introductory Readings in Geographic Information Systems*. London, New York: Taylor & Francis. 30-51.
- Daniel, M.-P.; Denis, M. (1998): Spatial descriptions as navigation aids: A cognitive analysis of route directions. *Kognitionswissenschaft* 7, 45-52.
- Davis, E. (1981): *Organizing Spatial Knowledge*. Yale: Yale University, Dept. of Computer Science. = Research Paper 193.
- Day, D.L.; Kovacks, D.K. (Hrsg.) (1996): *Computers, Communication, and Mental Models*. London: Taylor & Francis.
- De Berg, M.; van Krefeld, M.; Overmars, M.; Schwarzkopf, O. (1997): *Computational Geometry. Algorithms and Applications*. Berlin: Springer.
- De Vooght, G.; Vandierendonck, A. (1998): Spatial mental models in linear reasoning. *Kognitionswissenschaft* 7, 5-10.
- DeLoache, J.S.; Burns, V.M. (1994): Early understanding of the representational function of pictures. *Cognition* 52, 83-110.
- DeLoache, J.S.; Pierroutsakos, S.L.; Uttal, D.H.; Rosengren, K.S.; Gottlieb, A. (1998): Grasping the nature of pictures. *Psychological Science* 9 (3), 205-210.
- Denis, M.; Cocude, M. (1989): Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology* 1, 293-307.
- Denis, M.; Cocude, M. (1992): Structural properties of visual images constructed from poorly or well structured verbal descriptions. *Memory & Cognition* 20, 497-506.
- Denis, M.; Goncalves, M.-R.; Memmi, D. (1995): Mental scanning of visual images generated from verbal descriptions: Towards a model of image accuracy. *Neuropsychologica* 33, 1511-1530.
- Densham, P.J. (1994): Integrating GIS and spatial modelling: Visual active modelling and location selection. *Geographical Systems* 1 (3), 203-219.
- Dey, T.K.; Edelsbrunner, H.; Guha, S. (1996): Computational topology. In: *Discrete and Computational Geometry - Ten Year later*. AMS-IMS-SIAM Joint Summer Research Conference, Mount Holyoke College, South Hadley, MA, July 13-19, 1996. South Hadley, MA.
- Didier, A.; Grange, O. (1995): A fast scan-line algorithm for topological filling of well-nested objects in 2.5D digital pictures. *Theoretical Computer Science* 147 (1-2), 211-248.
- Dieberger, A. (1995): Providing spatial navigation for the world wide web. In: Frank, A.U.; Kuhn, W. (Hrsg.) (1995): *Spatial Information Theory: A Theoretical Basis for GIS*. Proceedings of the International Conference COSIT '95, Semmering, Austria, 21.9.-23.9.1995. Berlin: Springer. = Lecture Notes in Computer Science 988. 93-106.
- Ding, Y.; Fotheringham, A.S. (1992): The integration of spatial analysis and GIS. *Computers, Environment and Urban Systems* 16, 3-19.

- Do, E.Y.-L.; Gross, M.D. (1995): Drawing Analogies: Finding visual references by scetching. In: Proceedings of the 1995 National Conference of the Association for Computer Aided Design in Architecture, Seattle. Seattle.
- Dobkin, D.; Silver, D. (1988): Recipes for geometry and numerical analysis. Part 1: an empirical study. In: Proceedings of the 4th Annual ACM Symposium on Computational Geometry, Urbana-Champaign, IL. New York: ACM Press. 93-105.
- Dollinger, F.; Strobl, J. (Hrsg.) (1989a): Angewandte Geographische Informationstechnologie. Beiträge zum GIS-Symposium 5.7.-7.7.1989, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 13.
- Dollinger, F.; Strobl, J. (1989b): GIS - Begriff und Bedeutung. In: Dollinger, F.; Strobl, J. (Hrsg.) (1989a): a.a.O. 7-10.
- Dollinger, F.; Strobl, J. (Hrsg.) (1993): Angewandte Geographische Informationsverarbeitung V. Beiträge zum GIS-Symposium, 7.7.-9.7.1993, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 20.
- Dollinger, F.; Strobl, J. (Hrsg.) (1994): Angewandte Geographische Informationsverarbeitung VI. Beiträge zum GIS-Symposium, 6.7.-8.7.1994, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 21.
- Dollinger, F.; Strobl, J. (Hrsg.) (1995): Angewandte Geographische Informationsverarbeitung VII. Beiträge zum GIS-Symposium, 5.7.-7.7.1995, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 22.
- Dollinger, F.; Strobl, J. (Hrsg.) (1996): Angewandte Geographische Informationsverarbeitung VIII. Beiträge zum GIS-Symposium, 3.7.-5.7.1996, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 24.
- Dollinger, F.; Strobl, J. (Hrsg.) (1997): Angewandte Geographische Informationsverarbeitung IX. Beiträge zum GIS-Symposium, 2.7.-4.7.1997, Salzburg. Salzburg: Institut für Geographie der Universität Salzburg. = Salzburger Geographische Materialien 26.
- Downs, R.M.; Stea, D. (Hrsg.) (1973): Image and Environment: Cognitive Mapping and Spatial Behavior. Chicago: Aldine.
- Downs, R.M.; Stea, D. (1977): Maps in Minds. New York: Harper & Row.
- Downs, R.M.; Stea, D. (1982): Kognitive Karten: Die Welt in unseren Köpfen. New York: Harper & Row. = Uni-Taschenbücher 1126.
- Du, D.-Z.; Hwang, F.K. (Hrsg.) (1995): Computing in Euclidean Geometry. Singapore: World Scientific.
- Duchan, J.F.; Bruder, G.A.; Hewitt, L.E. (Hrsg.) (1995): Deixis in Narrative: A Cognitive Science Perspective. Hillsdale, NJ.: Erlbaum.
- Dumfarth, E. (1996): Interpolation von Oberflächen mit CRS (completely regularized spline with smoothing and tension). Evaluierung von Methode und Ergebnisse am Beispiel 'Bodenpreisoberflächen für die Stadt Salzburg'. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 27-38.
- Eco, U.; Santambrogio, M.; Violi, P. (Hrsg.) (1988): Meaning and Mental Representations. Bloomington: Indiana University Press.
- Edamura, T. (1997): Computers and Urban Planning. European Journal of Cognitive Psychology 2 (2), 151-153.
- Edelsbrunner, H.; Mücke, E.P. (1990): Simulation of simplicity: A technique to cope with degenerate cases in geometric algorithms. ACM Transactions on Graphics 9, 66-104.
- Egenhofer, M.J.; Frank, A.U. (1990): Object-oriented modelling in GIS: Inheritance and propogation. Auto-Carto (1990), 9, 588-598.
- Egenhofer, M.J.; Mark, D.M. (1995): Naive Geography. In: Frank, A.U.; Kuhn, W. (Hrsg.) (1995): Spatial Information Theory: A Theoretical Basis for GIS. Proceedings of the International Conference COSIT '95, Semmering, Austria, 21.9.-23.9.1995. Berlin: Springer. = Lecture Notes in Computer Science 988. 1-15.
- Ehrlich, K.; Johnson-Laird, P.N. (1982): Spatial descriptions and referential continuity. Journal of Verbal Learning and Verbal Behavior 21, 296-306.

- Eilan, N.; McCarthy, R.A.; Brewer, B. (Hrsg.) (1993): *Spatial Representation: Problems in Philosophy and Psychology*. Oxford: Blackwell.
- Emiris, I.Z.; Canny, J.F. (1995): A general approach to removing degeneracies. *SIAM Journal of Computing* 24, 650-664.
- Emmorey, K.; Kosslyn, S.M.; Bellugi, U. (1993): Visual imagery and visio-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition* 46, 139-181.
- Falmagne, R.J. (Hrsg.) (1975): *Reasoning: Representation and Process in Children and Adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Farah, M.J. (1985): Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology* 114, 91-103.
- Farah, M.J. (1988a): Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review* 95, 307-317.
- Farah, M.J. (1988b): The neuropsychology of mental imagery: Converging evidence from brain-damaged and normal subjects. In: Stiles-Davis, J.; Kritchesky, M.; Bellugi, U. (Hrsg.) (1988) a.a.O. 33-56.
- Farah, M.J.; Hammond, K.M.; Levine, D.N.; Calvanio, R. (1988a): Visual and spatial mental imagery: Dissociable systems of representations. *Cognitive Psychology* 20, 439-462.
- Farah, M.J.; Peronnet, F.; Gonon, M.A.; Girard, M.H. (1988b): Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology* 117, 248-257.
- Fenk, A. (1994): Spatial metaphors and logical pictures. In: Schnotz, W.; Kulhavi, R.W. (Hrsg.) (1994): *Comprehensions on Graphics*. Amsterdam: North Holland. = *Advances in Psychology* 108. 43-62.
- Ferguson, E.L.; Hegarty, M. (1994): Properties of cognitive maps constructed from texts. *Memory & Cognition* 22 (4), 455-473.
- Fischbein, E. (1987): *Intuition in Science and Mathematics*. Dordrecht: Reidel. = *Mathematical Education Library* 5.
- Fischer, M.M. (1994): Expert systems and artificial neural networks for spatial analysis and modelling. *Geographical Systems* 1 (3), 221-235.
- Fischer, M.M.; Nijkamp, P. (1993a): Design and use of geographic information systems and spatial models. In: Fischer, M.M.; Nijkamp, P. (Hrsg.) (1993b): a.a.O. 1-13.
- Fischer, M.M.; Nijkamp, P. (Hrsg.) (1993b): *Geographic Information Systems, Spatial Modelling, and Policy Evaluation*. Berlin: Springer.
- Fischer, M.M.; Scholten, H.J.; Unwin, D.J. (Hrsg.) (1996): *Spatial Analytical Perspectives on GIS*. Taylor & Francis. = *GISData* 4.
- Foley, J.; van Dam, A. (1984): *Fundamentals of Computer Graphics*. Reading, Mass.: Addison-Wesley.
- Foley, J.; van Dam, A.; Feiner, S.K.; Hughes, J.F. (1990): *Computer Graphics: Principles and Practice*. Reading, Mass.: Addison-Wesley. = *Systems Programming Series*.
- Fortune, S.J. (1993): Progress in computational geometry. In: Martin, R. (Hrsg.) (1993): *Directions in Geometric Computing*. Winchester: Information Geometers. 81-128.
- Fortune, S.J.; Milenkovic, V.J. (1991): Numerical stability of algorithms for line arrangements. In: *Proceedings of the 7th Annual ACM Symposium on Computational Geometry*, North Conway, CA. New York: ACM Press. 334-341.
- Fotheringham, A.S.; Densham, P.J.; Curtis, A. (1995): The zone definition problem in location-allocation modelling. *Geographical Analysis* 27, 61-77.
- Fotheringham, A.S.; Rogerson, P.A. (Hrsg.) (1992): *GIS and Spatial Analysis: Report on the Specialist Meeting*. Santa Barbara, CA: National Center for Geographic Information and Analysis (NCGIA). = *Report* 92-11.
- Fotheringham, A.S.; Rogerson, P.A. (1993): GIS and spatial analytical problems. *International Journal of Geographic Information Systems* 7 (1), 3-19.
- Fotheringham, A.S.; Rogerson, P.A. (1994a): Spatial analysis and GIS. *Geographical Systems* 1 (3), 175-177.

- Fotheringham, A.S.; Rogerson, P.A. (Hrsg.) (1994b): *Spatial Analysis and GIS*. London, New York: Taylor & Francis.
- Fotheringham, A.S.; Wong, D.W.S. (1991): The modifiable area unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025-1044.
- Frank, A.U. (1996): Der Nutzen und der Preis von geographischer Information. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 61-70.
- Frank, A.U.; Campari, I. (Hrsg.) (1993): *Spatial Information Theory: A Theoretical Basis for GIS*. Berlin: Springer. = *Lecture Notes in Computer Science* 716.
- Frank, A.U.; Campari, I.; Formentini, U. (Hrsg.) (1992): *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. Berlin: Springer. = *Lecture Notes in Computer Science* 639.
- Frank, A.U.; Kuhn, W. (Hrsg.) (1995): *Spatial Information Theory: A Theoretical Basis for GIS*. Proceedings of the International Conference COSIT '95, Semmering, Austria, 21.9.-23.9.1995. Berlin: Springer. = *Lecture Notes in Computer Science* 988.
- Frank, A.U.; Mark, D.M. (1991): Language issues for GIS. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991): *Geographical Information Systems. Principles and Applications*. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical. 147-163.
- Franklin, W.R.; Venkateshkumar, S.; Sun, D.; Kankanhalli, M.; Narayanaswami, C. (1994): Calculating the area of overlaid polygons without constructing the overlay. *Cartography and Geographic Information Systems*, 1-20.
- Freie Hansestadt Bremen, Der Senator für Bau, Verkehr und Stadtentwicklung (1998): *Stadtentwicklungskonzept Bremen – Entwurf*. Bremen.
- Freska, C.; Habel, C.; Wender, K.F. (Hrsg.) (1998): *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Berlin: Springer. = *Lecture Notes in Artificial Intelligence* 1404.
- Fudos, I.; Hoffmann, C. (1997): A graph-constructive approach to solving systems of geometric constraints. *ACM Transactions on Graphics* 16 (2), 179-216.
- Gärling, T.; Böök, A.; Lindberg, E. (1979): The acquisition and use of an internal representation of the spatial layout of the environment during locomotion. *Man-Environment Systems* 9, 200-208.
- Gärling, T.; Golledge, R.G. (Hrsg.) (1993): *Behavior and Environment: Psychological and Geographical Approaches*. Amsterdam: North Holland. = *Advances in Psychology* 96.
- Garner, B.J. (1993): GIS for everybody - everyday. *GIS Europe* 2 (2), 9-10.
- Garnham, A. (1992): Minimalism versus constructivism: A false dichotomy in theories of inference in reading. *Psychology* [on-line-serial].
- Garnham, A. (1997): Representing information in mental models. In: Conway, M.A. (Hrsg.) (1997): *Cognitive Models of Memory*. Hove: Psychology Press. = *Studies in Cognition*. 149-172.
- Gathercole, S.E. (1997): Models of verbal short-term memory. In: Conway, M.A. (Hrsg.) (1997): *Cognitive Models of Memory*. Hove: Psychology Press. = *Studies in Cognition*. 13-45.
- Gazzaniga, M.S. (1995): *The Cognitive Neurosciences*. Cambridge, MA.: MIT Press.
- Gehlke, C.E.; Biehl, K.K. (1934): Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association* (Suppl. 29), 169-170.
- Genter, D.; Stevens, A.L. (1983): *Mental Models*. Hillsdale: Erlbaum.
- Gerber, R.; Kwan, T. (1994): A phenomenographical approach to the study of pre-adolescent's use of maps in a wayfinding exercise in a suburban environment. *Journal of Environmental Psychology* 14, 265-280.
- Glasgow, J.; Narayanan, N.H.; Chandrasekaran, B. (Hrsg.) (1995): *Diagrammatic Reasoning. Cognitive and Computational Perspectives*. Menlo Park, CA: AAAI Press / The MIT Press.
- Glassner, A. (1994a): Clipping a Concave Polygon. In: Paeth, A. (Hrsg.) (1994): a.a.O. 50-54.
- Glassner, A. (Hrsg.) (1994b): *Graphics Gems I*. Boston: Academic Press.
- Glenberg, A.M. (1997): What memory is for. *Behavioral & Brain Sciences* 20, 1-55.

- Glenberg, A.M.; Langston, W.E. (1992): Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory & Language* 31, 129-151.
- Golledge, R.G.; Bell, S.; Dougherty, V. (1994): The cognitive map as an internal GIS. Paper presented at the meeting of the American Association of Geographers. San Francisco, CA.
- Golledge, R.G.; Hubert, L.J. (1982): Some comments on non-Euclidean mental maps. *Environment and Planning A* 14, 107-118.
- Golledge, R.G.; Ruggles, A.J.; Pellegrino, J.W.; Gale, N.D. (1993): Integrating route knowledge in an unfamiliar neighborhood. *Journal of Environmental Psychology* 13, 293-307.
- Golledge, R.G.; Stimson, R.J. (1997): *Spatial behavior: A geographic perspective*. New York, London: Guilford Press.
- Gombrich, E. (1990): Pictorial Instructions. In: Barlow, H.; Blakemore, C.; Weston-Smith, M. (Hrsg.) (1990) a.a.O. 26-45.
- Goodchild, M.F. (1990): Geographical data modelling. Santa Barbara, CA: National Center for Geographic Information and Analysis (NCGIA). = NCGIA Technical Report 90-11.
- Goodchild, M.F. (1995): GIS and geographic research. In: Pickles, J. (Hrsg.) (1995) a.a.O. 31-50.
- Goodchild, M.F.; Parks, B.O.; Steyaert, L.T. (Hrsg.) (1993): *Environmental Modelling with GIS*. New York, Oxford: Oxford University Press.
- Goodman, J.E.; O'Rourke, J. (Hrsg.) (1997): *Handbook of Discrete and Computational Geometry*. Boca Raton, New York: CRC Press. = CRC Press Series on Discrete Mathematics and its Applications.
- Goodman, N. (1990): Pictures in the mind? In: Barlow, H.; Blakemore, C.; Weston-Smith, M. (Hrsg.) (1990): *Images and Understanding. Thoughts about Images. Ideas about Understanding*. Cambridge, New York: Cambridge University Press. 358-364.
- Göpfert, W. (1991): *Raumbezogene Informationssysteme*. Wichmann.
- Graham, S. (1998): The end of geography or the explosion of place? Conceptualizing space, place, and information technology. *Progress in Human Geography* 22 (2), 165-185.
- Greene, D.H.; Yao, F.F. (1986): Finite-resolution computational geometry. In: *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Sciences*. Washington: IEEE Computer Society Press. 143-152.
- Guyote, M.J.; Sternberg, R.J. (1981): A transitive chain theory of syllogistic reasoning. *Cognitive Psychology* 13, 461-525.
- Gyselinck, V.; Tardieu, H. (1994): Illustrations, mental models, and comprehension of instructional text. In: Schnotz, W.; Kulhavi, R.W. (Hrsg.) (1994): *Comprehensions on Graphics*. Amsterdam: North Holland. = *Advances in Psychology* 108.139-151.
- Hagget, P. (Hrsg.) (1973): *Einführung in die Kultur- und sozialgeographische Regionalanalyse*. Berlin, New York: de Gruyter.
- Hanhwe, K.; Hirtle, S.C. (1995): Spatial metaphors and disorientation in hypertext browsing. *Behavior and Information Technology* 14 (4), 239-250.
- Hard, G. (1970): Was ist eine Landschaft? Über Ethymologie als Denkform in der geographischen Literatur. In: Bartels, D. (Hrsg.) (1970): *Wirtschafts- und Sozialgeographie*. Köln, Berlin: Kiepenheuer & Witsch. = *Neue Wissenschaftliche Bibliothek* 35. 66-84.
- Hard, G. (1982a): Landschaft. In: Jander, L.; Schramke, W.; Wenzel, H.-J. (Hrsg.) (1982) a.a.O. 160-170.
- Hard, G. (1982b): Geodeterminismus/Umweltdeterminismus. In: Jander, L.; Schramke, W.; Wenzel, H.-J. (Hrsg.) (1982) a.a.O. 104-106.
- Hard, G. (1982c): Länderkunde. In: Jander, L.; Schramke, W.; Wenzel, H.-J. (Hrsg.) (1982) a.a.O. 144-160.
- Hard, G. (1986): Der Raum - einmal systemtheoretisch gesehen. *Geographica Helvetica* (1986), 2, 77-83.
- Hard, G. (1990): 'Was ist Geographie?'. Reflektionen über geographische Reflexionstheorien. Karlsruhe: Institut für Geographie und Geoökologie der Universität Karlsruhe. = *Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie* 94.

- Harris, B.; Batty, M. (1993): Locational models, geographic information and planning support systems. *Journal of Planning Education and Research* 12, 184-198.
- Hatfield, G. (1990): *The Natural and the Normative. Theories of Spatial Perception from Kant to Helmholtz.* Cambridge, MA: MIT Press.
- Haubner, K. (1982): Was ist eigentlich Raumordnungspolitik? Begründungen, Zielsetzungen, Konzeptionen. In: Landeszentrale für politische Bildung (Hrsg.) (1982): *Raumordnungspolitik.* Stuttgart. = *Bürger im Staat* 1057. 9-20.
- Heckbert, P. (Hrsg.) (1994): *Graphics Gems IV.* Boston: Academic Press.
- Hellige, J.B.; Michimata, C. (1989): Categorization versus distance: Hemispheric differences for processing spatial informations. *Memory & Cognition* 17, 770-776.
- Henle, M. (1962): The relation between logic and thinking. *Psychological Review* 69, 366-378.
- Hernandez, D. (1994): *Qualitative Representation of Spatial Knowledge.* Berlin: Springer. = *Lecture Notes in Computer Science* 804.
- Hettner, A. (1927): *Die Geographie, ihre Geschichte, ihr Wesen und ihre Methoden.* Breslau.
- Hickok, G.; Sey, K.; Bellugi, U.; Klima, E.S. (1996): The basis of hemispheric asymmetry for language and spatial cognition: Clues from focal brain damage in two deaf native signers. *Aphasiology* 10 (6), 577-591.
- Hilgemeier, M. (1994): Teaming up - distributed applications in GIS. *GIS Europe* 3 (2), 16-16.
- Hirtle, S.C.; Jonides, J. (1985): Evidence of hierarchies in cognitive maps. *Memory & Cognition* 13, 208-217.
- Hoffmann, C. (1989): The problems of accuracy and robustness in geometric computation. *IEEE Computer* 22 (3), 31-41.
- Hoffmann, C.; Hopcroft, J.; Karasick, M. (1988): Towards implementing robust geometric computations. In: *Proceedings of the 4th Annual ACM Symposium on Computational Geometry, Urbana-Champaign, IL.* New York: ACM Press. 106-117.
- Holdings, C.S. (1994): Further evidence for the hierarchical representation of spatial information. *Journal of Environmental Psychology* 14 (2), 137-147.
- Huttenlocher, J.; Newcombe, N.; Sandberg, E.H. (1994): The coding of spatial location in young children. *Cognitive Psychology* 27, 115-147.
- Inhelder, B.; Piaget, J. (1958): *The Growth of Logical Thinking from Childhood to Adolescence.* London: Routledge & Kegan Paul.
- Jackendoff, R. (1987): On beyond zebra: The relation of linguistic and visual information. *Cognition* 26, 89-114.
- Jackendoff, R. (Hrsg.) (1992): *Languages of the Mind. Essays on Mental Representation.* Cambridge, MA.: MIT Press.
- Jackendoff, R.; Landau, B. (1991): Spatial language and spatial cognition. In: Napoli, D.J.; Kegl, J. (Hrsg.) (1991): *Bridges between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman.* Hillsdale, NJ: Erlbaum.
- Jander, L.; Schramke, W.; Wenzel, H.-J. (Hrsg.) (1982): *Metzler Handbuch für den Geographieunterricht.* Stuttgart.
- Jesorsky, C. (1991): *Komponenten eines Geographischen Informationssystems für die Landschafts- und Umweltplanung.* Berlin: TU Berlin. = *Landschaftsentwicklung und Umweltforschung* 79.
- Johnson, M. (1987): *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press.
- Johnson, M.; Lakoff, G. (1982): *Metaphors and Communication.* Trier. = *Reproduced: Universität Trier, Linguistic Agency. Paper. Series A* 97.
- Johnson-Laird, P.N. (1975): Models of deduction. In: Falmagne, R.J. (Hrsg.) (1975): *Reasoning: Representation and Process in Children and Adults.* Hillsdale, NJ: Lawrence Erlbaum Associates.
- Johnson-Laird, P.N. (1980): Mental models in cognitive science. *Cognitive Science* 4, 71-115.
- Johnson-Laird, P.N. (1983): *Mental Models. Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, MA.: Harvard University Press.

- Johnson-Laird, P.N. (1985): Deductive reasoning ability. In: Sternberg, R.J. (Hrsg.) (1985): *Human Abilities. An Information Processing Approach*. New York: Freeman. 173-194.
- Johnson-Laird, P.N. (1992): Mental models. In: Shapiro, S.C. (Hrsg.) (1992): *Encyclopedia of Artificial Intelligence*. New York, NY: Wiley. 932-939.
- Johnson-Laird, P.N. (1994): Mental models and probabilistic thinking. *Cognition* 50, 189-209.
- Johnson-Laird, P.N.; Byrne, R.M.J. (1991): *Deduction*. Hove: Lawrence Erlbaum Associates. = *Essays in Cognitive Psychology*.
- Johnson-Laird, P.N.; Byrne, R.M.J.; Tabossi, P. (1989): Reasoning by model: The case of multiple quantification. *Psychological Review* 96, 658-673.
- Johnson-Laird, P.N.; Steedman, M.J. (1978): The psychology of syllogisms. *Cognitive Psychology* 6, 372-375.
- Johnston, R.J. (1980): *Multivariate Statistical Analysis in Geography*. Harlow: Longman. = Longman Scientific & Technical.
- Johnston, R.J. (1994): On spatial analysis, place, and realism. *Urban Geography* 15 (3), 290-295.
- Karasick, M.; Strip, D. (1995): Intersecting solids on a massively parallel processor. *ACM Transactions on Graphics* 14 (1), 21-57.
- Keating, E. (1995): Spatial conceptualizations of social hierarchy in Pohnpei, Micronesia. In: Frank, A.U.; Kuhn, W. (Hrsg.) (1995): *Spatial Information Theory: A Theoretical Basis for GIS. Proceedings of the International Conference COSIT '95, Semmering, Austria, 21.9.-23.9.1995*. Berlin: Springer. = *Lecture Notes in Computer Science* 988. 463-474.
- Kilchenmann, A. (1991): *Geographie / Geoökologie und GIS*. In: Schilcher, M. (Hrsg.) (1991): *Geo-Informatik: Anwendungen, Erfahrungen, Tendenzen. Beiträge zum Internationalen Anwenderforum „Geoinformatik und Umweltinformatik“ in Duisburg, 20.-21.2.1991*. Berlin, München: Siemens AG. 127-134.
- Kirby, K.N.; Kosslyn, S.M. (1990): Thinking visually. *Mind and Language* 5, 324-341.
- Kirk, D. (Hrsg.) (1994): *Graphics Gems III*. Boston: Academic Press.
- Kitchin, R.M. (1993): Bringing psychology and geography closer. *Journal of Environmental Psychology* 13, 183-186.
- Kitchin, R.M. (1994a): Cognitive maps: What are they and why study them? *Journal of Environmental Psychology* 14, 1-19.
- Kitchin, R.M. (1994b): Semiotics, marxism, and spatial conception: A new paradigm? *Journal of Environmental Psychology* 14, 83-85.
- Kitchin, R.M. (1994c): Spatial familiarity as a variable in cognitive mapping. *Swansea Geographer* 31, 42-52.
- Kitchin, R.M. (1996): Exploring approaches to computer cartography and spatial analysis in cognitive mapping research: CMAP and MiniGASP prototype packages. *Cartographic Journal* 33 (1), 51-55.
- Kitchin, R.M. (1997a): Exploring spatial thought. *Environment and Behavior* 29 (1), 123-156.
- Kitchin, R.M. (1997b): Relations between psychology and geography. *Environment and Behavior* 29 (4), 554-573.
- Klauer, K.C. (1998): Visuelle und räumliche Interferenzen beim Schlußfolgern. *Kognitionswissenschaft* 7 (1), 11-18.
- Klosterman, R.E.; Xie, Y. (1997): Retail Impact Analysis with loosely coupled GIS and a spreadsheet. *International Planning Studies* 2 (2), 175-192.
- Klüter, H. (1986): *Raum als Element sozialer Kommunikation*. Gießen. = *Gießener Geographische Schriften* 60.
- Klüter, H. (1987): Räumliche Orientierung als sozialgeographischer Grundbegriff. *Geographische Zeitschrift* 75 (2), 86-98.
- Klüter, H. (1994): *Raum als Objekt menschlicher Wahrnehmung und Raum als Element sozialer Kommunikation*. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 136, 143-178.
- Knauff, M.; Rauh, R.; Schlieder, C. (1995): Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In: *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ.: Lawrence Erlbaum Associates. 200-205.



- Knauff, M.; Rauh, R.; Schlieder, C.; Strube, G. (1998): Mental models in spatial reasoning. In: Freska, C.; Habel, C.; Wender, K.F. (Hrsg.) (1998): *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Berlin: Springer. = *Lecture Notes in Artificial Intelligence* 1404, 267-291.
- Koch, R. (1992): Informationssysteme für die räumliche Planung und ihre Grenzen durch Statistik und Datenschutz. Eine Zwischenbilanz aus bayerischer Sicht. *Raumforschung und Raumordnung* (1992), 3-4, 136-141.
- Kohonen, T. (1982): Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, 59-69.
- Kosslyn, S.M. (1980): *Image and Mind*. Cambridge, MA.: Harvard University Press.
- Kosslyn, S.M. (1987): Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review* 94, 148-175.
- Kosslyn, S.M. (1994): *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA.: MIT Press.
- Kosslyn, S.M. (Hrsg.) (1995a): *Frontiers in Cognitive Neuroscience*. Cambridge, Mass.: MIT Press.
- Kosslyn, S.M. (1995b): *Visual Cognition*. Cambridge, Mass.: MIT Press. = Osherson, P.N.: *An Invitation to Cognitive Science*, Vol. 2.
- Kosslyn, S.M.; Koenig, O.; Barret, A.; Cave, C.B.; Tang, J.; Gabrieli, J.D.E. (1989): Evidence for two types of spatial representation: Hemispheric specialization for categorial and coordinate spatial relations. *Journal of Experimental Psychology: Human Perception & Performance* 15, 723-735.
- Krishnan, S.; Manocha, D. (1997): An efficient surface intersection algorithm based on lower dimensional formulation. *ACM Transactions on Graphics* 16 (1), 74-106.
- Kriz, J. (1981): *Methodenkritik empirischer Sozialforschung. Eine Problemanalyse sozialwissenschaftlicher Forschungspraxis*. Stuttgart: Teubner. = *Studienskripte zur Soziologie* 49.
- Kriz, K.; Riedl, A. (1994): Kartographische Visualisierung in einem Informationssystem. In: Dollinger, F.; Strobl, J. (Hrsg.) (1994) a.a.O. 391-397.
- Kuhn, W. (1993): Various perspectives on interface metaphors. In: EGIS Foundation (Hrsg.) (1993): *EGIS '93. Fourth European Conference and Exhibition on Geographical Information Systems*, Genoa, Italy, 29.3.-1.4.1993 (Vol. 1). Utrecht, Amsterdam: EGIS Foundation. 456-463.
- Kuhn, W.; Frank, A.U. (1991): A formalization of metaphors and image-schemas in user interfaces. In: Mark, D.M.; Frank, A.U. (Hrsg.) (1991): *Cognitive and Linguistic aspects of geographic space*. Dordrecht: Kluwer Academic Publishers. 419-434.
- Kuipers, B. (1978): Modelling spatial knowledge. *Cognitive Science* 2, 129-153.
- Kulpa, Z. (1994): Diagrammatic representation and reasoning. *Machine Graphics & Vision* 3 (1/2), 77-103.
- Kulpa, Z. (1997): Diagrammatic representation for a space of intervals. *Machine Graphics & Vision* 6 (1), 5-24.
- Lakoff, G. (1987): *Women, Fire, and Dangerous Things. What Categories Tell us about the Nature of Thought*. Chicago: University Press of Chicago.
- Lakoff, G. (1988): Cognitive semantics. In: Eco, U.; Santambrogio, M.; Violi, P. (Hrsg.) (1988): *Meaning and Mental Representations*. Bloomington: Indiana University Press. 119-154.
- Lakoff, G. (1990): The invariance hypothesis: Is Abstract reasoning based on image schemas? *Cognitive Linguistics* 1 (1), 39-74.
- Landau, B. (1986): Early map use as an unlearned ability. *Cognition* 22, 201-223.
- Landau, B. (1991): Spatial knowledge of objects in the young blind child. *Cognition* 38, 145-178.
- Landau, B.; Gleitman, L.R. (1985): *Language and Experience: Evidence from the Blind Child*. Cambridge, MA.: Harvard University Press.
- Landau, B.; Jackendoff, R. (1993): 'What' and 'Where' in spatial language and spatial cognition. *Behavioral & Brain Sciences* 16, 217-265.
- Landau, B.; Spelke, E.; Gleitman, H. (1984): Spatial knowledge in a young blind child. *Cognition* 16, 225-260.
- Langston, W.; Kramer, D.C.; Glenberg, A.M. (1998): The representation of space in mental models derived from text. *Memory & Cognition* 26 (2), 247-262.

- Larkin, J.H.; Simon, H.A. (1987): Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11, 65-99.
- Laszlo, M.J. (1995): *Computational Geometry and Computer Graphics in C++*. Upper Saddle River, NJ: Prentice Hall.
- Laurini, R.; Thompson, D. (1995): *Fundamentals of Spatial Information Systems*. London, San Diego: Academic Press (4. Auflage). = The A.P.I.C. Series 37.
- Levinson, S.C. (1996): Language and space. *Annual Review of Anthropology* 25, 353-382.
- Lewis, H.R.; Papadimitriou, C.H. (1981): *Elements of the Theory of Computation*. Englewood Cliffs, NJ: Prentice Hall. = Prentice-Hall Software Series.
- Leykauf, J.; Margraf, O. (1990): Zur Nutzung digitaler Technologien in der Geographie. *Petermanns Geographische Mitteilungen* (1990), 2, 101-105.
- Leyton, M. (1989): Inferring causal history from shape. *Cognitive Science* 13, 357-389.
- Liang, Y.; Barsky, B. (1983): An analysis and algorithm for polygon clipping. *Communications of the ACM* 26 (11), 868-877.
- Lloyd, P.E.; Dicken, P. (1972): *Location in Space: A Theoretical Approach to Economic Geography*. New York: Harper & Row.
- Lloyd, R. (1982): A look at images. *Annals of the Association of American Geographers* 72, 532-548.
- Lloyd, R. (1993): Cognitive processes and cartographic maps. In: Gärling, T.; Golledge, R.G. (Hrsg.) (1993): *Behavior and Environment: Psychological and Geographical Approaches*. Amsterdam: North Holland. = *Advances in Psychology* 96.
- Lloyd, R. (1994): Learning spatial prototypes. *Annals of the Association of American Geographers* 84, 418-440.
- Lloyd, R.; Patton, D.; Cammack, R. (1996): Basic level geographic categories. *The Professional Geographer* 48, 181-194.
- Logie, R.H. (1986): Visuo-spatial processes in working memory. *Quarterly Journal of Experimental Psychology* 38a, 229-247.
- MacEachren, A.M. (1990): Approaches to truth in geographic visualization. In: *Twelfth International Symposium on Computer-Assisted Cartography*, Charlotte, Carolina (Vol. 4). 110-118.
- MacEachren, A.M. (1995): *How Maps Work. Representation, Visualization, and Design*. New York, London: Guilford Press.
- MacLennan, B.J. (1993): Visualizing the possibilities. *Behavioral & Brain Sciences* 16, 356-357.
- Maguire, D.J. (1991): An Overview and Definition of GIS. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991): *Geographical Information Systems. Principles and Applications*. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical. 9-20.
- Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991a): *Geographical Information Systems. Principles and Applications*. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical.
- Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (1991b): Introduction. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991a): a.a.O. 3-7.
- Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (1991c): Introduction. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991a): a.a.O. 111-117.
- Maillot, P.-G. (1992): A new, fast method for 2D polygon clipping: Analysis and software implementation. *ACM Transactions on Graphics* 11 (3), 276-290.
- Majure, J.J.; Cressie, N. (1993): Explore: Exploratory spatial analysis with ARC/INFO. In: *Proceedings of the Thirteenth Annual ESRI User Conference*, Palm Springs, CA (Vol. 1). 247-277.
- Mandl, P. (1993): GIS-Modellierung und integrierte geographische Datenverarbeitung als Hauptarbeitsbereiche der GIS-Applikationsforschung in den 90er Jahren. In: Dollinger, F.; Strobl, J. (Hrsg.) (1993) a.a.O. 7-20.
- Mandl, P.; Wastl-Walter, D. (Hrsg.) (1990): *Theorie und Praxis geographischer Informationssysteme in Geographie und Raumplanung*. Wien: Arbeitskreis für Regionalforschung. = AMR Info 21 (4-6).

- Mani, K.; Johnson-Laird, P.N. (1982): The mental representations of spatial descriptions. *Memory & Cognition* 10, 181-187.
- Manocha, D.; Demmel, J. (1994): Algorithms for intersecting parametric and algebraic curves I: simple intersection. *ACM Transactions on Graphics* 13 (1), 73-100.
- Marchese, F.T. (Hrsg.) (1995): *Understanding Images*. New York: Springer.
- Margraf, O. (1994a): Zum Ansatz eines länderkundlichen GIS für Sachsen. In: Dollinger, F.; Strobl, J. (Hrsg.) (1994): a.a.O. 487-498.
- Margraf, O. (1994b): Zum Aufbau eines länderkundlichen GIS für Sachsen. In: Aurada, K. (Hrsg.) (1994): Beiträge des 10. Kolloquiums für Theorie und quantitative Methoden in der Geographie, Göhren auf Rügen, 23.2.-26.2.1994. Greifswald: Institut für Geographie der Ernst-Moritz-Arndt-Universität Greifswald. = Greifswalder Geographische Arbeiten 11. 161-174.
- Mark, D.M. (1989): Cognitive image schemata for geographic information systems. *GIS/LIS* 89 2, 551-560.
- Mark, D.M.; Frank, A.U. (1990): *Experiential and Formal Models of Geographic Space*. Santa Barbara, CA: National Center for Geographic Information and Analysis (NCGIA). = NCGIA Technical Report 90-10.
- Mark, D.M.; Frank, A.U. (Hrsg.) (1991): *Cognitive and Linguistic aspects of geographic space*. Dordrecht: Kluwer Academic Publishers.
- Mark, D.M.; Gould, M.D. (1995): Wayfinding directions as discourse: Verbal directions in English and Spain. In: Duchan, J.F.; Bruder, G.A.; Hewitt, L.E. (Hrsg.) (1995): *Deixis in Narrative: A Cognitive Science Perspective*. Hillsdale, NJ.: Erlbaum. 387-405.
- Marquardt-Kuron, A.; Schliephake, K. (Hrsg.) (1996): *Raumbezogene Verkehrswissenschaften - Anwendung mit Konzept*. Bonn: Irene Kuron. = Material zur Angewandten Geographie 26.
- Marr, D. (1982): *Vision*. San Francisco: Freeman.
- Martin, D. (1991): Representing the socioeconomic world. *Papers in Regional Science* 70 (3), 317-327.
- Martin, R. (Hrsg.) (1993): *Directions in Geometric Computing*. Winchester: Information Geometers.
- May, M. (1995): Diagrammatisches Denken: Zur Deutung logischer Diagramme als Vorstellungsschemata bei Lakoff und Peirce. *Zeitschrift für Semiotik* 11 (3-4), 285-305.
- McDermott, D. (1981): Finding objects with given spatial properties. Yale: Yale University, Dept. of Computer Science. = Research Paper 195.
- McHaffie, P.H. (1995): Manufacturing metaphors. Public cartography, the market, and democracy. In: Pickles, J. (Hrsg.) (1995) a.a.O. 113-129.
- McNamara, T.P. (1986): Mental representations of spatial relations. *Cognitive Psychology* 18, 87-121.
- McNamara, T.P.; Altarriba, J.; Bendele, M.; Johnson, S.C.; Clayton, K.N. (1989): Constraints on priming in spatial memory: Naturally versus experimental learned environments. *Memory & Cognition* 17 (4), 444-453.
- McNamara, T.P.; Hardy, J.K.; Hirtle, S.C. (1989): Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15, 211-227.
- McNamara, T.P.; Ratcliff, R.; McKoon, G. (1984): The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 4, 723-732.
- Medyckyj-Scott, D.; Blades, M. (1992): Human spatial cognition: Its relevance to the design and use of Spatial information systems. *Geoforum* 23 (2), 215-226.
- Mevenkamp, N. (1996a): Eine multivariate Visualisierungstechnik und ihre Anwendung zur explorativen Überprüfung formaler Typisierungen. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 212-217.
- Mevenkamp, N. (1996b): Geometrische Informationsverarbeitung - Ein GIS-Konzept zur Integration von (explorativer) Datenanalyse und thematischer Kartographie. In: Margraf, O. (Hrsg.) (1996): *Theorie und Quantitative Methodik in der Geographie*. Tagungsband 11. Deutschsprachiges Kolloquium des Arbeitskreises Leipzig/Naunhof, 28.2.-2.3.1996. Leipzig: Institut für Länderkunde Leipzig. = Beiträge zur Regionalen Geographie 42. 36-45.
- Milenkovic, V.J. (1989a): Double precision geometry: a general technique for calculating line and segment intersections using rounded arithmetic. In: *Proceedings of the 30th Annual Symposium of Computer Science (FOCS), IEEE, October 1989*. IEEE. 500-506.

- Milenkovic, V.J. (1989b): Robust geometric computations for vision and robotics. In: Proceedings of the DARPA Image Understanding Workshop, May 1989. DARPA. 764-773.
- Millar, S. (1994): Understanding and Representing Space. Theory and Evidence from Studies with Blind and Sighted Children. Oxford: Clarendon Press.
- Moll, P. (Hrsg.) (1995): Raumbezogene Informationssysteme in der Anwendung. Bonn: Irene Kuron. = Material zur Angewandten Geographie 23.
- Montani, C.; Re, M. (1987): Vector and raster hidden surface removal using parallel connected stripes. *IEEE Computer* 7 (7), 14-23.
- Montello, D.R. (1993): Scale and multiple psychologies of space. In: Frank, A.U.; Campari, I. (Hrsg.) (1993): *Spatial Information Theory: A Theoretical Basis for GIS*. Berlin: Springer. = *Lecture Notes in Computer Science* 716. 312-321.
- Montello, D.R.; Freudenschuh, S.M. (1995): Sources of spatial knowledge and their implications for GIS: An Introduction. *Geographical Systems* 2, 169-176.
- Morill, R.L. (1974): *The Spatial Organization of Society*. North Scituate, Mass.: Duxbury Press.
- Morrison, A. (1996): Public transport maps in Western Europe cities. *Cartographic Journal* 33 (2), 93-110.
- Morrow, L.; Ratcliff, G. (1988): Neuropsychology of spatial cognition: Evidence from cerebral lesions. In: Stiles-Davis, J.; Kritchesvky, M.; Bellugi, U. (Hrsg.) (1988) a.a.O. 5-32.
- Morton, J. (1967): Interaction of information in word recognition. *Nature* 215, 203-204.
- Mount, D.M. (1997): Geometric Intersection. In: Goodman, J.E.; O'Rourke, J. (Hrsg.) (1997): *Handbook of Discrete and Computational Geometry*. Boca Raton, New York: CRC Press. = *CRC Press Series on Discrete Mathematics and its Applications*. 615-630.
- Murdock, B.B.Jr. (1965): Effects of subsidiary tasks on short-term memory. *British Journal of Psychology* 56, 413-419.
- Napoli, D.J.; Kegl, J. (Hrsg.) (1991): *Bridges between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*. Hillsdale, NJ: Erlbaum.
- Nelson, E.S. (1996): A cognitive map experiment: mental representations and the encoding process. *Cartography and Geographic Information Systems* 23 (4), 229-248.
- Neprash, J.A. (1934): Some problems in the correlation of spatially distributed variables. *Journal of the American Statistical Association* (Suppl. 29), 167-169.
- Neumann, J. von (1958): *The Computer and the brain*. New Haven, Conn.
- Newcombe, N.; Liben, L.S. (1982): Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology* 34, 46-58.
- Nickerson, R.S.; Adams, M.J. (1979): Long-term memory for a common object. *Cognitive Psychology* 11, 287-307.
- Nijkamp, P.; Scholten, H.J. (1993): Spatial information systems: design, modelling, and use in planning. *International Journal of Geographic Information Systems* 7 (1), 85-96.
- Nyerges, T.L. (1991): Geographic information abstractions: Conceptual clarity for geographic modeling. *Environment and Planning A* 23, 1483-1499.
- Nyerges, T.L.; Mark, D.M.; Laurini, R.; Egenhofer, M.J. (Hrsg.) (1995): *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*. Dordrecht: Kluwer. = *NATO ASI Series D: Behavioral and Social Sciences* 83.
- O'Keefe, J. (1991): The hippocampal cognitive map and navigational strategies. In: Paillard, J. (Hrsg.) (1991): *Brain and Space*. Oxford, New York, Tokyo: Oxford University Press. 379-403.
- O'Keefe, J.; Nadel, L. (1978): *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
- O'Rourke, J. (1994): *Computational Geometry in C*. New York: Cambridge University Press.
- O'Rourke, J.; Badler, N. (1980): Image analysis of human motion constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 522-536.
- Oberlander, J. (1998): Abstraction and individual style in syllogistic and heterogeneous reasoning. *Kognitionswissenschaft* 7, 33-38.

- Olson, D.R.; Bialystock, W. (1983): *Spatial Cognition. The Structure and Development of Mental Representations of Spatial Relations*. Hillsdale, NJ.: Erlbaum.
- Openshaw, S. (1978): An empirical study of some zone-design criteria. *Environment and Planning A* 10, 710-794.
- Openshaw, S. (1990): Spatial analysis and GIS: A review of progress and possibilities. In: Scholten, H.J.; Stillwell, J.C.H. (Hrsg.) (1990): *Geographical Information Systems for Urban and Regional Planning*. Dordrecht: Kluwer.
- Openshaw, S. (1991): Developing appropriate spatial analysis methods for GIS. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991): *Geographical Information Systems. Principles and Applications*. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical. 389-402.
- Openshaw, S. (1994): Two exploratory space-time-attribute pattern analysers relevant to GIS. In: Fotheringham, A.S.; Rogerson, P.A. (Hrsg.) (1994): *Spatial Analysis and GIS*. London, New York: Taylor & Francis.
- Openshaw, S. (1995): A review of GIS in business applications. *Geographical Systems* 2 (2), 153-168.
- Openshaw, S.; Cross, A.; Charlton, M. (1990): Building a prototype geographical correlates exploration machine. *International Journal of Geographic Information Systems* 4 (3), 297-311.
- Openshaw, S.; Scholten, H.J. (1993): *Spatial Analysis and Geographic Information Systems: a Challenge for the 1990's*. Utrecht, Amsterdam: EGIS Foundation. = Manuskript zum Workshop Spatial Analysis. EGIS '93.
- Osherson, D. (1975): Logic and models of logical thinking. In: Falmagne, R.J. (Hrsg.) (1975): *Reasoning: Representation and Process in Children and Adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ottmann, T.; Thiemt, G.; Ullrich, C. (1987): Numerical stability of geometric algorithms. In: *Proceedings of the 3rd Annual ACM Symposium on Computational Geometry*. New York: ACM Press. 119-125.
- Ottmann, T.; Widmayer, P. (1993): *Algorithmen und Datenstrukturen*. Mannheim, Leipzig, Wien, Zürich: BI-Wissenschaftsverlag. = Reihe Informatik 70.
- Ottoson, T. (1988): What does it take to read a map? *Cartographica* 25, 28-35.
- Paeth, A. (Hrsg.) (1994): *Graphics Gems V*. Boston: Academic Press.
- Paillard, J. (Hrsg.) (1991): *Brain and Space*. Oxford, New York, Tokyo: Oxford University Press.
- Paivio, A. (1969): Mental imagery in associative learning and memory. *Psychological Review* 76, 241-263.
- Paivio, A. (1971): *Imagery and verbal processes*. New York: Holt.
- Paivio, A. (1986): *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.
- Patrikalakis, N.M.; Cho, W.; Hu, C.-Y.; Maekawa, T.; Sherbrooke, E.C.; Zhou, J. (1995): Towards robust geometric modellers. 1994 progress report. In: *Proceedings of the 1995 National Conference of the Association for Computer Aided Design in Architecture*, Seattle. Seattle. 139-140.
- Payne, D.; Conrad, F. (Hrsg.) (1996): *Intersections in Basic and Applied Memory Research*. Mahwah, NJ.: Erlbaum.
- Peuquet, D.J.; Marble, D.F. (Hrsg.) (1990): *Introductory Readings in Geographic Information Systems*. London, New York: Taylor & Francis.
- Piaget, J. (1954): *The construction of reality in the child*. New York: Basic Books.
- Piaget, J.; Inhelder, B. (1967): *The Child's Conception of Space*. New York: Norton.
- Pick, H.; Acredolo, L.P. (1983): *Spatial Orientation: Theory, Research; and Application*. New York, NY: Plenum Press.
- Pickering, A. (1992): From science as knowledge to science as practice. In: Pickering, A. (Hrsg.) (1992): *Science as Practice and Culture*. Chicago, London: University of Chicago Press. 1-26.
- Pickles, J. (1985): *Phenomenology, Science and Geography. Spatiality and the Human Sciences*. Cambridge, Ma.: Cambridge University Press. = Cambridge Human Geography.
- Pickles, J. (Hrsg.) (1995): *Ground Truth. The Social Implications of Geographic Information Systems*. New York, London: Guilford Press. = Mappings: Society, Theory, Space.
- Poincaré, H. (1913): *The Foundation of Science*. Lancaster, P.A.: Science Press.

- Portugali, J. (Hrsg.) (1996): *The Construction of Cognitive Maps*. Dordrecht: Kluwer. = *The GeoJournal Library* 32.
- Postman, L.; Phillips, L.W. (1965): Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology* 17, 140-150.
- Ponn, A. (1996): Salzburger Straßenverkehrsinformationssystem. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 24.80-84.
- Prasad, M. (1994): Intersection of Line Segments. In: Arvo, J. (Hrsg.) (1994): *Graphics Gems II*. Boston: Academic Press. 7-9.
- Pylyshyn, Z.W. (1973): What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80, 1-24.
- Pylyshyn, Z.W. (1979): The rate of 'mental rotation' of images: A test of a holistic analogue hypothesis. *Memory & Cognition* 7, 19-28.
- Pylyshyn, Z.W. (1981): The imagery debate: Analogue media versus tacit knowledge. *Psychological Review* 86, 16-45.
- Pylyshyn, Z.W. (1984): *Computation and Cognition. Toward a Foundation of Cognitive Science*. Cambridge, MA: MIT Press.
- Rase, W.D. (1996): Interpolation von stetigen Oberflächen aus flächenbezogenen Informationen. In: Dollinger, F.; Strobl, J. (Hrsg.) (1996) a.a.O. 17-26.
- Rase, W.D.; Sinz, M. (1993): Kartographische Visualisierung von Planungskonzepten. *Kartographische Nachrichten* 1993 (4) 139-145.
- Rauh, R.; Knauff, M.; Schlieder, C.; Strube, G. (1998): Leitthema: Räumliche mentale Modelle / Spatial Mental Models. *Kognitionswissenschaft* 7, 1-4.
- Rauh, R.; Schlieder, C.; Knauff, M. (1997): Präferierte mentale Modelle beim räumlich relationalen Schließen: Empirie und kognitive Modellierung. *Kognitionswissenschaft* 6, 21-34.
- Regian, J.W.; Yadrick, R.M. (1994): Assessment of configurational knowledge of naturally- and artificially-acquired large-scale space. *Journal of Environmental Psychology* 14, 211-223.
- Rips, L.J. (1983): Cognitive processes in propositional reasoning. *Psychological Review* 90, 38-71.
- Ritter, H.; Kohonen, T. (1989): Self-organizing semantic maps. *Biological Cybernetics* 61, 241-254.
- Roberts, S.M.; Schein, R.H. (1995): Earth shattering. Global imagery and GIS. In: Pickles, J. (Hrsg.) (1995) a.a.O. 171-195.
- Robinson, W.S. (1950): Ecological correlation and the behavior of individuals. *American Sociological Review* 15, 351-357.
- Rodrigue, J.-P. (1995): The heuristic classification of functional land use: A knowledge-based approach. *Geographical Systems* 2 (2), 103-120.
- Rosch, E.; Mervis, C. (1975): Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7, 573-605.
- Rosch, E.; Mervis, C.; Gray, C.; Johnson, D.; Boyes-Braem, P. (1976): Basic objects in natural categories. *Cognitive Psychology* 8, 382-439.
- Rothkegel, R.; Wender, K.F.; Schumacher, S. (1998): Judging spatial relations from memory. In: Freska, C.; Habel, C.; Wender, K.F. (Hrsg.) (1998): *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Berlin: Springer. = *Lecture Notes in Artificial Intelligence* 1404. 79-105.
- Rutland, A.; Custance, D.; Campbell, R.N. (1993): The ability of three- to four-year-old children to use a map in a large-scale environment. *Journal of Environmental Psychology* 13, 365-372.
- Rybash, J.M.; Hoyer, W.J. (1992): Hemispheric specialization for categorical and coordinate spatial representations: A reappraisal. *Memory & Cognition* 20 (3), 271-276.
- Sachs, J.S. (1967): Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2, 437-442.

- Salamé, P.; Baddeley, A.D. (1982): Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior* 21, 150-161.
- Salamé, P.; Baddeley, A.D. (1987): Noise, unattended speech and short-term memory. *Ergonomics* 30, 1185-1193.
- Sanocki, T.; Epstein, W. (1997): Priming spatial layout of scenes. *Psychological Science* 8 (5), 374-378.
- Sassen, S. (1997): Electronic Space: Embedded and Segmented. *International Planning Studies* 2 (2), 229-240.
- Scaife, M.; Rogers, Y. (1996): External cognition: How do graphical representations work? *International Journal of Human-Computer Studies* 45, 185-213.
- Schaeken, W.; Giroto, V.; Johnson-Laird, P.N. (1998): The effect of an irrelevant premise on temporal and spatial reasoning. *Kognitionswissenschaft* 7, 27-32.
- Schaeken, W.; Johnson-Laird, P.N.; d'Ydewalle, G. (1996): Mental models in temporal reasoning. *Cognition* 60, 205-234.
- Schilcher, M. (Hrsg.) (1991): *Geo-Informatik: Anwendungen, Erfahrungen, Tendenzen. Beiträge zum Internationalen Anwenderforum „Geoinformatik und Umweltinformatik“ in Duisburg, 20.-21.2.1991.* Berlin, München: Siemens AG.
- Schilcher, M. (1996): Geoinformationssysteme - Zwischenbilanz einer stürmischen Entwicklung. *Zeitschrift für Vermessungswesen* 121 (8), 363-377.
- Schlieder, C. (1996): Diagrammatic reasoning about Allen's interval relations. In: *AAAI Spring Symposium on Cognitive and Computational Models of Spatial Representations.* Stanford University, CA., 25.3.-27.3.1996. Stanford. = *AAAI Technical Report SS-96-03.*
- Schlieder, C. (1997): Diagrammatic transformation processes on relational maps. In: Anderson, M. (Hrsg.) (1997): *Reasoning with diagrammatic representations.* AAAI Fall Symposium on Reasoning with Diagrammatic Representations, MIT, Cambridge, Mass., 8.11.-10.11.1997. Menlo Park, CA: AAAI Press. = *AAAI Technical Report 97-03.*
- Schmithüsen, J. (1948): „Fliesengefüge der Landschaft“ und „Ökotoptop“. *Vorschläge zur begrifflichen Ordnung und zur Nomenklatur in der Landschaftsforschung.* *Berichte zur deutschen Landeskunde* 5, 74-83.
- Schnotz, W. (1992): *Wissenserwerb mit logischen Bildern.* Tübingen: Deutsches Institut für Fernstudien an der Universität Tübingen. = *Forschungsberichte* 58.
- Schnotz, W.; Kulhavi, R.W. (Hrsg.) (1994): *Comprehensions on Graphics.* Amsterdam: North Holland. = *Advances in Psychology* 108.
- Schnotz, W.; Picard, E.; Henninger, M. (1994): The use of graphics and texts in constructing mental models. In: Schnotz, W.; Kulhavi, R.W. (Hrsg.) (1994): *Comprehensions on Graphics.* Amsterdam: North Holland. = *Advances in Psychology* 108.185-205.
- Scholten, H.J.; Stillwell, J.C.H. (Hrsg.) (1990): *Geographical Information Systems for Urban and Regional Planning.* Dordrecht: Kluwer.
- Schorn, P. (1994): Degeneracy in geometric computation and the perturbation approach. *Computer Journal* 37, 35-42.
- Schramke, W. (1975): *Zur Paradigmengeschichte der Geographie und ihrer Didaktik. Eine Untersuchung über Geltungsanspruch und Identitätskrise eines Faches.* Göttingen: Selbstverlag. = *Geographische Hochschulmanuskripte* 2.
- Schulz, H.-D. (1980): *Die deutschsprachige Geographie von 1800-1970. Ein Beitrag zur Geschichte ihrer Methodologie.* Berlin. = *Abhandlungen des Geographischen Instituts, Anthropogeographie* 29.
- Sechrest, S.; Greenberg, D. (1981): A visible polygon reconstruction algorithm. *Computer Graphics* 15 (3), 17-26.
- Segal, M.G.; Sequin, C.H. (1985): Consistent calculation for solids modelling. In: *Proceedings of the 1st Annual ACM Symposium on Computational Geometry.* New York: ACM Press. 29-38.
- Shallice, T.; Warrington, E.K. (1970): Independent functioning of verbal memory stores; A neuropsychological study. *Quarterly Journal of Experimental Psychology* 22, 261-273.

- Shepard, R. (1990): On understanding mental images. In: Barlow, H.; Blakemore, C.; Weston-Smith, M. (Hrsg.) (1990): *Images and Understanding. Thoughts about Images. Ideas about Understanding*. Cambridge, New York: Cambridge University Press. 365-370.
- Shepard, R.N.; Metzler, J. (1971): Mental rotation of three-dimensional objects. *Science* 171, 701-703.
- Shneiderman, B. (1992): Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics* 11 (1), 92-99.
- Slovan, A. (1971): Interactions between philosophy and A.I. The role of intuition in intelligence. *Artificial Intelligence* 2, 209-225.
- Spiekermann, K.; Wegener, M. (1993): Zeitkarten für die Raumplanung. *Informationen zur Raumentwicklung* (1993), 7, 459-487.
- Staab, S.; Hahn, U. (1997): Dimensional reasoning with qualitative and quantitative distances. *Proc. Of the AAAI-Workshop on Spatial and Temporal Reasoning*, Providence, RI, USA, 1997.
- Stea, D.; Blaut, J.M.; Stephens, J. (1996): Mapping as a cultural universal. In: Portugali, J. (Hrsg.) (1996): *The Construction of Cognitive Maps*. Dordrecht: Kluwer. = *The GeoJournal Library* 32. 345-360.
- Stenning, K. (1998): Aligning logical and psychological contributions to the understanding of human reasoning. *Kognitionswissenschaft* 7 (1), 39-44.
- Stenning, K.; Oberlander, J. (1995): A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science* 19, 97-140.
- Stephan, E.-M. (1990): Exploratory data visualization for reliable integration. In: *Twelfth International Symposium on Computer-Assisted Cartography*, Charlotte, Carolina (Vol. 4). 100-109.
- Sternberg, R.J. (Hrsg.) (1985): *Human Abilities. An Information Processing Approach*. New York: Freeman.
- Sternberg, R.J. (1990): *Metaphors of Mind: Conception of the Nature of Intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (Hrsg.) (1994): *Thinking and Problem Solving. Handbook of Perception and Cognition*. San Diego, CA: Academic Press. = 2. Ed.
- Stevens, A.; Coupe, P. (1978): Distortions in judged spatial relations. *Cognitive Psychology* 10, 422-437.
- Stiles-Davis, J.; Kitchesvky, M.; Bellugi, U. (Hrsg.) (1988): *Spatial Cognition. Brain Bases and Development*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Storkebaum, W. (Hrsg.) (1967): *Zum Gegenstand und zur Methode der Geographie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Sugihara, K. (1989): On finite-precision representations of geometric objects. *Journal of Computers and Systems Sciences* 39, 236-246.
- Sutherland, E.E.; Hodgeman, G.W. (1974): Reentrant polygon clipping. *Communications of the ACM* 17 (1), 32-42.
- Taylor, H.A.; Tversky, B. (1995): Assessing spatial representation using text. *Geographical Systems* 2, 235-254.
- Taylor, H.A.; Tversky, B. (1996): Perspective in spatial descriptions. *Journal of Memory & Language* 35, 371-391.
- Taylor, P.J.; Johnston, R.J. (1995): Geographic information systems and geography. In: Pickles, J. (Hrsg.) (1995) a.a.O. 51-67.
- Thorndyke, P.W. (1981): Distance estimation from cognitive maps. *Cognitive Psychology* 13, 526-550.
- Tlanka, M.; Wilson, P.N. (1994): The effect of landmarks on route learning in a computer-simulated environment. *Journal of Environmental Psychology* 14, 305-313.
- Tobler, W.R. (1970): A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234-240.
- Tolman, E.C. (1948): Cognitive maps in rats and men. *Psychological Review* 55, 189-208.
- Tomlin, C.D. (1990): *Geographic Information Systems and Cartographic Modelling*. Englewood Cliffs, NJ: Prentice Hall.



- Tomlin, C.D. (1991): Cartographic modelling. In: Maguire, D.J.; Goodchild, M.F.; Rhind, D.W. (Hrsg.) (1991): Geographical Information Systems. Principles and Applications. Vol. 1: Principles. Burnt Hill, Harlow: Longman Scientific and Technical. 361-374.
- Tomlinson, R.F. (1990): Geographic Information Systems - a new frontier. In: Peuquet, D.J.; Marble, D.F. (Hrsg.) (1990): Introductory Readings in Geographic Information Systems. London, New York: Taylor & Francis. 18-29.
- Troll, C. (1950): Die geographische Landschaft und ihre Erforschung. *Studium Generale* 3 (4/5), 163-181.
- Tufte, E.R. (1983): *The Visual Display of Quantitative Information*. Cheshire, Conn.: Graphics Press.
- Tukey, J.W. (1977): *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.
- Tulving, E. (1966): Subjective Organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 5, 193-197.
- Turing, A.M. (1937): On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Second Series* 42, 230-265.
- Tversky, B. (1993): Cognitive maps, cognitive collages, and spatial mental models. In: Frank, A.U.; Campari, I. (Hrsg.) (1993): *Spatial Information Theory: A Theoretical Basis for GIS*. Berlin: Springer. = Lecture Notes in Computer Science 716. 14-24.
- Tversky, B. (1995): Cognitive origins of graphic conventions. In: Marchese, F.T. (Hrsg.) (1995): *Understanding Images*. New York: Springer. 29-53.
- Tversky, B. (1996): Memory for pictures, environments, maps, and graphs. In: Payne, D.; Conrad, F. (Hrsg.) (1996): *Intersections in Basic and Applied Memory Research*. Mahwah, NJ.: Erlbaum. 257-277.
- Ullman, S. (1984): Visual routines. *Cognition* 18, 97-159.
- Unwin, D.J. (1996): GIS, spatial analysis and spatial statistics. *Progress in Human Geography* 20 (4), 540-551.
- Usery, E. (1993): Category theory and the structure of features in geographic information systems. *Cartography and Geographic Information Systems* 20, 5-12.
- van Teeffelen, P.; van Grunsven, L.; Verkoren, O. (Hrsg.) (1992): Possibilities and Constraints of GIS Applications in Developing Countries. *Proceedings of the GISDECO Seminar Organized by the GIS Working Group of the Faculty of Geographical Sciences, Utrecht University, 7.4.-8.4.1991*. Utrecht: Rijksuniversiteit Utrecht. = *Nederlandse Geografische Studies* 152.
- Vatti, B.R. (1992): A generic solution to polygon clipping. *Communications of the ACM* 35 (7), 56-63.
- Venn, J. (1880): On the diagrammatic and mechanical representation of propositions and reasoning. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 9, 1-18.
- Veregin, H. (1995): Computer innovation and adoption in geography. A critique of conventional technological models. In: Pickles, J. (Hrsg.) (1995) a.a.O. 88-112.
- Vogel, M.; Blaschke, T. (1996): GIS in Naturschutz und Landschaftspflege: Überblick über Wissensstand, Anwendungen und Defizite. *Laufener Seminarbeiträge* (1996), 4, 7-19.
- von Frieling, H.-D.; Güssefeldt, J. (1996): Eine GIS-Anwendung zum Test der Theorie städtischer Bodenpreise und Einwohnerverteilung. In: Margraf, O. (Hrsg.) (1996): *Theorie und Quantitative Methodik in der Geographie. Tagungsband 11. Deutschsprachiges Kolloquium des Arbeitskreises Leipzig/Naunhof, 28.2.-2.3.1996*. Leipzig: Institut für Länderkunde Leipzig. = *Beiträge zur Regionalen Geographie* 42. 46-59.
- Wagener, M.; Wender, K.F. (1985): Spatial representations and inference processes in memory for text. In: Rickheit, G.; Strohner, H. (Hrsg.) (1985): *Inferences in Text Processing*. Amsterdam: Kluwer. 115-136.
- Walmsley, D.J.; Lewis, G.J. (1993): *People and Environment: Behavioral Approaches in Human Geography*. London: Longman Scientific and Technical (2. Aufl.).
- Walton, D.J.; Xu, R. (1991): Turning point preserving planar interpolation. *ACM Transactions on Graphics* 10 (3), 297-311.
- Weber, E.S. (1986): Systems to think with: A response to 'A vision for decision support systems'. *Journal of Management Information Systems* 2 (4), 85-97.
- Webster, C.J. (1994): GIS and the scientific inputs to planning. Part 2: prediction and prescription. *Environment and Planning B: Planning and Design* 21, 145-157.

- Weiler, K.; Atherson, P. (1974): Hidden surface removal using polygon area sorting. Proceedings of SIGGRAPH 11 (2), 214-222.
- Weizenbaum, J. (1976): Die Macht der Computer und die Ohnmacht der Vernunft. Frankfurt am Main: Suhrkamp. = suhrkamp taschenbuch wissenschaft 274.
- Wender, K.F.; Wagener, M. (1990): Zur Verarbeitung räumlicher Information: Modelle und Experimente. Kognitionswissenschaft 1, 4-14.
- White, R.; Engelen, G. (1994): Cellular dynamics and GIS: Modelling spatial complexity. Geographical Systems 1 (3), 237-253.
- Wilton, R.N. (1979): Knowledge of spatial relations: The specification of the information used in making inferences. Quarterly Journal of Experimental Psychology 31, 133-146.
- Wittgenstein, L. (1947): Bemerkungen über die Philosophie der Psychologie. Band 1. Frankfurt a.M.: Suhrkamp. = Suhrkamp-Taschenbuch Wissenschaft 507.
- Worboys, M.F. (1995): GIS. A Computing Perspective. London: Taylor & Francis.
- Yap, C.K. (1997): Robust geometric computation. In: Goodman, J.E.; O'Rourke, J. (Hrsg.) (1997): Handbook of Discrete and Computational Geometry. Boca Raton, New York: CRC Press. = CRC Press Series on Discrete Mathematics and its Applications. 653-668.
- Yap, C.K.; Dube, T. (1995): The exact computation paradigm. In: Du, D.-Z.; Hwang, F.K. (Hrsg.) (1995): Computing in Euclidean Geometry. Singapore: World Scientific. 452-486.