

In Kombination mit den klassischen und etablierten Verfahren (Interviews, Beobachtungen, Textanalysen) kann sich ein möglichst vollständiges und gegenstandadäquates Bild der untersuchten Fälle ergeben.

## 9.5 Methodologischer Exkurs: Bild und Text

Bilder und Texte sind seit langem – und, was die Bilder betrifft, spätestens mit dem sogenannten *visual* oder *iconic turn* wieder verstärkt – Gegenstand der Methoden- und Theoriedebatte in den Geistes-, Kultur- und Sozialwissenschaften (z. B. Boehm 2004; Bräunlein 2004; Hartinger 2007; Breckner 2010). Wenn es um Bilder und Texte als Daten zu einem bestimmten Forschungsgebiet geht oder wenn der Status ihrer Existenzweisen bestimmt werden soll (»Was sind Bilder?«, »Was sind Texte?«), wird oft betont, dass Bilder und Texte grundverschieden seien.

In diesem methodologischen Exkurs möchte ich nicht bestreiten, dass Bilder und Texte verschiedene Dinge – genauer: verschiedene Datengattungen – sind. Aber sie sind, so meine Hypothese, auch nicht so verschieden, dass man sie nicht prinzipiell mit den gleichen Methoden untersuchen könnte. Dieses Argument werde ich im Folgenden ausführen und dabei weniger vorschlagen, dass »Bilder wie Texte sind« – weil dies reflexhaft (und manchmal begründet) die Kritik eines ›Textualismus‹ auslöst –, sondern umgekehrt zu zeigen versuchen, dass »Texte wie Bilder sind« bzw. Bildern ähnlicher als oft angenommen.

Um die angesprochene Debatte produktiv weiterführen zu können, gilt es eine zentrale, aber meines Wissens oft vernachlässigte Unterscheidung deutlich zu markieren: Die Unterscheidung zwischen *Texten und Bildern als Daten* in sozial-, kultur- und geisteswissenschaftlicher Forschung auf der einen Seite und »*Text*« und »*Bild*« als *theoretischen Konzepten* bzw. als epistemologischen Metaphern auf der anderen Seite. Dieser Unterschied wird in den betreffenden Diskussionen nicht immer deutlich markiert: Wenn man beispielsweise formuliert, dass »Bilder nicht wie Texte gelesen« werden sollen, dann sind »Bilder« in dieser Formulierung das untersuchte Datenmaterial. Der Begriff »Texte« in dieser Formulierung kann aber sowohl das textliche Datenmaterial meinen – dann vergleicht man zwei Datengattungen miteinander und kommt zu dem (wenig überraschenden) Schluss, dass sie verschieden sind –, als auch auf das theoretische Konzept »Text« (in welcher Variante auch immer) verweisen. Umgekehrt ist der Begriff »Bild« beispielsweise im Kompositum »Sprachbild« kein Verweis auf bildliches Datenmaterial, sondern ein theoretisches (metaphorisches) Konzept, um auf eine bestimmte Funktionsweise von Sprache hinzuweisen.

Es ist auffällig, dass Begriffe im semantischen Feld »Text« (z. B. »Bedeutung«, »Gewebe«, »Metapher«, »Semiotik«) als theoretische Konzepte deutlich stärker ausgearbeitet worden sind als Begriffe im semantischen Feld »Bild« (z. B. »Ikonogra-

fie«, was aber stärker deskriptiv angelegt ist). Dieser Umstand verweist auf ein Problem, das unten weiter ausgeführt wird. Kurz: »Bilder« werden oft als Dinge in der Welt beschrieben, denen bestimmte Eigenschaften zugewiesen werden; der Begriff »Text« hingegen kann sich mal auf empirisches Datenmaterial beziehen, mal auf ein mehr oder weniger komplex ausgearbeitetes theoretisches Konzept (so bspw. in Barthes bezeichnendem Aufsatz »From Work to Text«, 1979).

Um das Argument vorzubereiten, dass »Texte wie Bilder sind« bzw. den Bildern ähnlicher als oft angenommen, fasse ich zunächst einige häufig formulierte methodologische Positionen über Bilder und Texte *als Daten* historischer und sozialwissenschaftlich-empirischer Forschung zusammen. Diese Positionen zeigen – so meine These –, dass in der methodologisch teilweise sehr hart gezogenen Grenzlinie zwischen den Datengattungen Bild und Text längst überkommen geglaubte Dichotomien reproduziert werden – beispielsweise Natur/Kultur, Körper/Geist, Gefühl/Verstand – und das Bild dabei der ersten Seite der Unterscheidung (Natur, Körper, Gefühl) zugewiesen wird.

Außerdem fällt auf: Was *Bilder* sind, wird in der methodologischen Debatte in der Regel vor der impliziten oder expliziten Vergleichsfolie des *Textes* erörtert. Dabei wird vielfach vorausgesetzt, dass mehr oder weniger klar sei, was textliche Daten sind. Somit unterliegen, provokativ formuliert, bildwissenschaftliche Zugänge, die sich ausdrücklich von einem »Textualismus« anderer Zugänge abgrenzen, selbst einem *impliziten Textualismus*: Was ein Bild ist und wie es zu untersuchen sei, wird häufig aus der Perspektive des Textes (und im Medium des Textes) formuliert und das heißt in Abgrenzung zum Text. So schreibt Boehm beispielsweise:

»Bilder besitzen eine eigene, nur ihnen zugehörige Logik. Unter Logik verstehen wir: die konsistente Erzeugung von Sinn aus genuin bildnerischen Mitteln. Und erläuternd füge ich hinzu: Diese Logik ist nicht-prädikativ, das heißt nicht nach dem Muster des Satzes oder anderer Sprachformen gebildet. Sie wird nicht gesprochen, sie wird wahrnehmend realisiert« (Boehm 2004, 28–29).

Davon abgesehen, dass auch Boehm hier eine kognitive Textmetapher (»Logik« von λογική τέχνη bzw. λογικός = das Wort, die Vernunft oder das Denken betreffend) verwendet, erhebt er außerdem – mindestens implizit – den Anspruch, eine »nicht-prädikative« Logik des Bildes in das wissenschaftliche Medium des Textes zu überführen.

Wie auch andere Autor:innen im Feld der Bildwissenschaften geht Boehm davon aus, dass die Bedeutung von Bildern »nicht gesprochen«, sondern »wahrnehmend realisiert« wird. Dennoch ist das Ziel dieses Unterfangens, etwas über die Bedeutung von Bildern zu sagen und zu schreiben. Darin liegt eine methodologische und erkenntnistheoretische Herausforderung: Wie kann etwas, das wesentlich

nicht-prädikativ ist (die Logik des Bildes, in Boehms Terminologie) mit sprachlichen Mitteln, die einer prädikativen Logik folgen, ausgedrückt werden?

Man muss Boehm und anderen zustimmen, dass Bilder einen nicht-sprachlichen Wirkungsraum haben, eine Art von Bedeutung, die betrachtende Menschen »wahrnehmend« realisieren, ohne dafür Worte finden zu müssen. Doch gibt es begründete Zweifel daran, dass dieses unmittelbare Erleben bereits Gegenstand einer sozialwissenschaftlichen Untersuchung sein kann, die sich eben auf das Soziale, also den kommunikativ realisierten Ausdruck individuellen Erlebens bezieht.

Nicht zuletzt haben auch Texte diese Art von nicht-prädikativer, unmittelbar affektiver Wirkung: Ein Gefühl oder eine Stimmung, die lesende Menschen überkommt, ohne dass sie dies recht in Worte fassen könnten oder müssten. Ohne jede Art von sprachlicher Logik können Texte affektiv berühren. Doch diese unmittelbare, affektive Wirkung von Texten wird erst dann Gegenstand sozialwissenschaftlicher Untersuchung, wenn sie als sozialer Sachverhalt – und das heißt im Wesentlichen: als kommunikativer Sachverhalt – auftritt (Krech 2021, 99).

Daher schlage ich vor, eine übergeordnete Perspektive zu finden, die Bilder und Texte als Daten versteht, die zwar verschieden sind, aber doch ähnlich genug, um im Rahmen eines gemeinsamen methodologischen Paradigmas untersucht zu werden: nämlich als semiotische Gefüge, das heißt als Arrangement von Elementen, die in ihren je spezifischen empirisch und historisch vorfindlichen Konfigurationen Bedeutung ermöglichen und realisieren. So macht Edmund Hermsen bereits 2003 den Vorschlag, dass man »mithilfe der Vorstellung, dass ikonische und anikonische Zeichen nur jeweils andere Modi der Repräsentation bilden [...] »unaufgereggt« die Konkurrenz von Text- und Bildcode auflösen« könne (Hermsen 2003, 118). Dieser Vorschlag wurde in der Literatur aber noch nicht genügend verfolgt.

Wie angedeutet gibt es eine Reihe von methodologisch viel beachteten Positionen, die davon ausgehen, dass Bilder einer prinzipiell anderen Logik folgen als Texte (und Sprache) und daher nicht mit Methoden untersucht werden können, die für Textanalysen entwickelt wurden. Dazu gehört auch der bereits zitierte Ansatz von Boehm, demzufolge Bilder eine »eigene, nur ihnen zugehörige Logik« besitzen (Boehm 2004, 28–29).

In ähnlicher Weise fordert Hans Belting, »Bilder nicht mehr mit Texten zu erklären, sondern von Texten zu unterscheiden« (2001, 15). Dieser auch von William Mitchell (1994) vertretene Ansatz wird unter anderem in der Religionswissenschaft rezipiert und zugleich als Gegenentwurf zu einer semiotischen Herangehensweise behandelt: »Den semiotischen Ansatz, der gekennzeichnet ist durch das Bemühen, eine gemeinsame Grundlage von Text und Bild über eine Zeichentheorie zu erstellen, die wiederum an das Modell ›Sprache‹ angelehnt ist, lässt Mitchell hinter sich« (Bräunlein 2004, 205). Was Bräunlein hier über den semiotischen Ansatz schreibt: dass er eine gemeinsame, zeichentheoretisch fundierte Grundlage für die Analyse von Text und Bild formulieren will, teile ich. Lediglich darüber, wie eine solche Zei-

chentheorie aussehen könnte und wie sie Sprache – und welche Art von Sprache? – als »Modell« benutzt, besteht meiner Ansicht nach Diskussionsbedarf. Wenn man eine Einführung der Semiotik auf Text und Sprache hinter sich lässt, kann die Semiotik auch als theoretischer Rahmen für die Untersuchung von Bildern und anderen Medien dienen. So betont Umberto Eco bereits in den 1970er Jahren, dass eine semiotische Perspektive auf Kultur nicht zwangsläufig Kultur insgesamt auf Text und Sprache reduziert:

»[...] to reduce the whole of culture to semiotics does not mean that one has to reduce the whole of material life to pure mental events. To look at the whole of culture *sub specie semiotica* is not to say that culture is only communication and signification but that it can be understood more thoroughly if it is seen from the semiotic point of view. And that objects, behavior and relationships of production and value function as such socially precisely because they obey semiotic laws« (Eco 1976, 27).

Aber was ist – in der fächerübergreifenden Literatur, die sich mit der Analyse von Bildern befasst – ein »Bild«, welche Eigenschaften hat es, und wie soll es demzufolge untersucht und ausgewertet werden? Bilder unterscheiden sich, folgt man der Literatur, in den folgenden Punkten von Texten:

**Synchronizität vs. Diachronizität** Bilder, so die häufig geäußerte Vermutung, seien synchron und nichtlinear (für die Geschichtswissenschaft z. B. Lengwiler 2011, 133–136) – im Gegensatz zu Texten, die diachron und linear seien. Statt von »synchron« sprechen manche Autor:innen von »simultan« bzw. von der »Simultanstruktur« des Bildes (Imdahl 1996, 23). Mit anderen Worten: Im Bild ist alles gleichzeitig und gemeinsam präsent, im Text folgen die Elemente dagegen linear aufeinander. Ralf Bohnsack versteht unter Simultaneität »das gleichzeitige Zusammentreffen nicht zusammenhängender Ereignisabläufe, also unterschiedlicher Sequenzen«, was nicht allein zeitlich (im Sinne von »gleichzeitig«) gemeint ist, sondern auch als »gemeinsam« im Sinne von holistisch. Synchronizität dagegen bezieht er auf die Zeitdimension, das Gleichzeitige des Bildes im Gegensatz zum Nacheinander von Sequenzen im Text (Bohnsack 2013, 88).

Dieses Verständnis von Bildern hat Konsequenzen für die Analyse: Ein »textförmiger Zugang zum Bild ist vor allem durch seine Sequenzialität, durch eine am zeitlichen Nacheinander orientierte Ordnung, charakterisiert« (Bohnsack 2013, 82). Unter Verweis auf Max Imdahl (1996) kritisiert Bohnsack, dass das Bild in der sprachlichen Repräsentation in ein narratives Nacheinander zergliedert werde. Im Bild sei aber alles simultan und synchron, »in Totalpräsenz« gegeben. Sequenzielle Interpretationsverfahren (wie beispielsweise sequenzanalytische Verfahren) seien daher nicht gegenstandsadäquat (Bohnsack 2013, 82).

Doch Bohnsack gesteht einen Aspekt der Linearität von Bildern ein, der erst im methodischen Vorgehen zum Tragen kommt, und zwar dann, wenn man sich den »Zugang zur Ganzheitlichkeit auf dem Wege der Simultaneität [...] erst erarbeiten [muss]. Dabei kann der Weg hin zu einer die Totalität des Bildes erfassenden Simultaneität durchaus sukzessiv (!) vonstattengehen« (Bohnsack 2013, 95). Der wesentliche methodische Unterschied bestehe dann aber darin, dass man Bilder aufgrund der Simultaneität ihrer Semantik immer vollständig analysieren müsse; die Untersuchung eines Ausschnitts (analog zu einer ausgewählten Passage aus einem Text oder Interviewprotokoll) reiche nicht aus, um das Bild im Ganzen zu verstehen. Das jedoch wäre empirisch zu überprüfen und unterscheidet sich vermutlich je nach untersuchtem Text oder Bild, nicht aber pauschal zwischen den Datensorten »Bilder« und »Texte«.

Der Aspekt der Simultaneität wird auch von Gottfried Boehm formuliert (z. B. Boehm 1987). Aus Perspektive der Rezeption eines Bildes betont er, dass es nicht um isolierte, aufeinander folgende Elemente geht, sondern um ein simultanes Feld. Hinsichtlich der Wahrnehmung von Bildern konzidiert Johannes Grave mit Verweis auf Boehm jedoch, dass »Bilder sowohl sukzessiv als auch simultan erfahren werden« (Grave 2014, 59–60). Er gesteht damit eine »rezeptionsästhetische Temporalität des Bildes« ein, die sich nicht in der Abfolge von wahrgenommenen Bildelementen erschöpft (Grave 2014, 59–60). Dennoch sei der »Akt des Betrachtens nicht analog zum Lesevorgang« zu verstehen, weil Bilder nicht aus »distinkten Grundeinheiten« wie Buchstaben, Silben oder Wörtern bestünden (Grave 2014, 63) (dies wird auch unter dem Aspekt der Nicht-Sprachlichkeit von Bildern ausgeführt, siehe unten). Wir haben damit einen zweiten Indikator dafür gegeben, dass Bilder – nicht nur in der methodischen Rekonstruktion (Bohnsack 2013, 95), sondern auch in ihrer Wahrnehmung durch Betrachter:innen (Grave 2014, 59–60) – einen linearen, diachronen Aspekt aufweisen.

Man könnte nun zu argumentieren versuchen, dass *bildliche Daten* linear und diachron sind (z. B., weil sie anders nicht produziert, betrachtet und kognitiv verarbeitet werden können). Es scheint mir aber interessanter zu zeigen, dass *textliche Daten* synchrone, simultane und nichtlineare Eigenschaften haben, also all jene Eigenschaften, die üblicherweise den Bildern zugeschrieben werden.

Ein alltagsweltlich einfaches Beispiel für die Simultaneität von Texten ist das Genre des Witzes. Ein Witz ist nur ein Witz, wenn man den Anfang vom Ende, der Pointe, her versteht. Ein Witz muss daher, auch wenn er diachron erzählt wird (anders ist es biologisch gesehen nicht möglich), in seiner Simultanstruktur erfasst werden, um alltagsweltlich verstanden und gegenstandsadäquat analysiert zu werden. Das gleiche gilt allgemeiner formuliert auch für andere Texttypen und Textteile: Sätze, Absätze, Kapitel und Bücher sind nur vollständig, weil der Anfang ein Ende hat. Auch wenn in einem Ausschnitt die Struktur des Ganzen angelegt ist, ist der Gegenstand erst in sich geschlossen, wenn der Anfang vom Ende her erfasst wird

und das Ende vom Anfang her. Daher haben textliche Daten einen rekursiven und simultanen Aspekt *und* einen linearen und diachronen Aspekt. Erkenntnistheoretisch basiert diese Feststellung auf der Annahme, dass sich »ein Potenzial immer nur *ex post*, nämlich anhand seiner Realisierung ausmachen lässt« (Krech 2021, 197)

Was die Rezeption des Textes – verstanden als Psychologie der Wahrnehmung – angeht, so zeigen psychologische Studien über das Lesen, dass das menschliche Auge nicht Buchstabe für Buchstabe, Silbe für Silbe und Wort für Wort linear verfolgt, sondern simultan – also gleichzeitig und gemeinsam – größere Einheiten erfasst, die als solche kognitiv verarbeitet werden (eine frühe Studie dieser Art stammt von Just und Carpenter 1980).

Denkt man schließlich an die medialen Aspekte eines Textes, das heißt die Art und Weise, wie er auf Papier, Leder, Papyrus, Leinwand etc. festgehalten und wie er gezeigt wird (auf Plakaten, in Büchern, in Zeitschriften, in digitalen Formaten etc.), so wird deutlich, dass Texte in ihrer medialen Verfasstheit synchron und simultan sind. Das Zeigen der Bibel als Teil der Liturgie ist dafür ein Beispiel: Der Text als Ganzes wird simultan und synchron präsentiert.

Kurzum: Die Unterscheidung »Bilder = synchron/simultan« vs. »Texte = diachron/linear« ist eine simplifizierende, die den Daten, die wir empirisch vorfinden, nicht gerecht wird. Beide Arten von Daten haben sowohl synchrone als auch diachrone Eigenschaften und die Methoden ihrer Auswertung müssen diesen Eigenschaften gerecht werden.

**Hohe Informationsdichte vs. niedrige Informationsdichte** Über Bilder wird oft geschrieben, sie würden auf verhältnismäßig kleinem Raum ungleich mehr Informationen unterbringen als Texte (z. B. Lengwiler 2011, 133–136). Karis fasst die vorherrschende Lehrmeinung präzise zusammen, wenn er schreibt:

»Ein Bild enthält grundsätzlich mehr Informationen, als in einem noch so ausführlichen Text wiedergegeben werden könnten. Bis in die Details der Farbigkeit, der Schattierungen und Formen hinein könnten Bilder beschrieben werden und noch immer würden sich aus einer Umsetzung dieses Beschreibungstextes in eine bildliche Darstellung höchst unterschiedliche Resultate ergeben. Während also ein Bild eine unendliche Zahl von Bildbeschreibungen produziert, die nie erschöpfend sein können, kann eine Bildbeschreibung eine beliebige Zahl von Bildern produzieren, die allesamt erschöpfend sind« (Karis 2013, 141).

Auch hier wird die Materialgattung »Bild« aus der Perspektive des Textes beschrieben. Dem Bild wird ein Informationsgehalt zugeschrieben, der so reichhaltig ist, dass er nie vollständig in Textform übertragen werden kann. Wenn Marcus Banks zu begründen versucht, warum Bildanalysen in den Sozialwissenschaften noch zu selten diskutiert würden, schreibt er sehr plakativ: »[V]isual information, visual data,

visual results, are simply too messy, too rich, too particular to be reduced to abstraction and linear theorization« (Banks 2014, 405). Hier schwingt zum einen die bereits angedeutete Dichotomie von Gefühl vs. Verstand und Irrationalität vs. Rationalität mit, aber auch die Informationsmenge (»rich«) wird betont, die es unmöglich mache, eine »lineare Theorie« auf der Basis von bildlichen Quellen zu entwickeln.

Die zunächst so einsichtige These über den textlich nicht fassbaren Informationsgehalt von Bildern ist meiner Ansicht nach zu befragen. Zum einen bleibt offen, was genau mit »Information« gemeint ist. Folgt man einem systemtheoretisch geprägten Kommunikationsbegriff, so ist Information das Ergebnis einer Selektion im Kommunikationsprozess (z. B. Luhmann und Baecker 2006, 128). Anders gesagt: Informationen sind nicht einfach da, zum Beispiel in einem Bild, und warten darauf, abgerufen, verstanden oder in Text überführt zu werden. Sie werden erst in einem kommunikativen Prozess konstituiert, wenn eine Auswahl der Information aus einer Mitteilung stattfindet.

Für Bilder kann das konkret bedeuten, dass sie ein großes Potenzial an möglichen Informationsselektionen beinhalten, aber nicht von sich aus bereits voller Informationen sind. Dies wiederum gilt genauso für Texte, womit die Texte den Bildern ähnlicher sind, als gemeinhin vermutet wird. Eine bestimmte Schattierung oder ein bestimmter Pinselstrich in einem Bild ist damit nicht *per se* eine Information, aber sie kann zur Information werden, wenn sie in einem kommunikativen Zusammenhang adressiert wird, der sich beispielsweise mit der Technik des Malers befasst. Ein detailreiches Landschaftsgemälde kann insofern *erschöpfend* in Text übersetzt werden, wenn der kommunikative Zusammenhang gerade damit befasst ist, Landschaftsmalerei von Portraitmalerei zu unterscheiden, weil es dann auf viele Details im Bild gar nicht ankommt. Umgekehrt wird man in Texten Merkmale wie die Schriftgröße, den Satzspiegel oder die Zeichensetzung nur dann als Information betrachten, wenn der kommunikative Zusammenhang dies erfordert. Ein Text enthält bis in die Etymologie der vorhandenen Wörter, ihre Grammatik, semantischen Assoziationen, phonetischen Eigenschaften hin ein »unendliches« Informationspotenzial, von dem aber im konkreten Kommunikationsprozess nur der jeweils relevante Teil realisiert und kommuniziert wird.

Nicht zuletzt steht die These vom überreichen Informationsgehalt der Bilder in einem auffälligen Kontrast zur häufig vehement vertretenen These, dass Bilder zuallererst nicht-sprachlich und affektiv wirksam seien (siehe unten) und eben nicht primär semantisch und damit potenziell informationshaltig.

Kurz: Die These von der unterschiedlichen Informationsdichte von bildlichen und textlichen Daten hinkt einerseits aufgrund des unklaren Informationsbegriffs, andererseits lassen sich zahlreiche Texte als Beispiele anführen, die gerade nicht »ordentlich« oder »eindeutig« sind, oder »erschöpfend« in ein Bild übersetzt werden könnten; die häufig als Mystik bezeichneten Traditionsstränge innerhalb und zwischen den religiösen Traditionen legen davon beredtes Zeugnis ab (man denke

beispielsweise an die jüdische Kabbala [z. B. Scholem 2001] oder die christliche Mystik [z. B. Heiler 1934]). Ganz davon abgesehen, dass eine gute Theorie nicht »linear« sein muss (was auch immer damit gemeint ist), sind alle von Banks (2014, 405) genannten Eigenschaften (»messy«, »rich«, »particular«) für Texte ebenso gültig wie für Bilder.

**Vieldeutigkeit vs. Eindeutigkeit** Eng verbunden mit dem Postulat der hohen Informationsdichte von Bildern gehen manche Autor:innen davon aus, dass Bilder im Gegensatz zu Texten vieldeutig seien. Die Vieldeutigkeit von Bildern – hier von Fernsehbildern – wird u. a. von Karis thematisiert, der, unter Bezug auf Hans-Georg Soeffner (2000), davon ausgeht, dass Bilder zumindest prinzipiell vieldeutiger sind als Texte bzw. Sprache. Auch sogenannter »Schlüsselbilder« seien zwar »als Produkte der Verdichtung Ergebnis eines Vereindeutigungsprozesses«, deshalb aber keineswegs eindeutig (Karis 2013, 140). So betont beispielsweise auch Bohnsack die Herausforderung, der »Vieldeutigkeit des Bildes« methodisch gerecht zu werden (Bohnsack 2013, 87). Noch stärker wird es aus bildlinguistischer Perspektive formuliert, wenn zum Beispiel Roman Opilowski schreibt, dass Bilder eine »offene Bedeutung« hätten, die im Text »konkretisiert« werden müsse (Opilowski 2011, 207).

Es mag für bestimmte Sorten von Texten stimmen, dass sie im pragmatischen Verwendungszusammenhang zumindest mehr oder weniger eindeutig erscheinen (beispielsweise Nachrichtentexte oder gut geschriebene Gebrauchs- und Bedienungsanleitungen), doch die überwiegende Mehrheit der Texte und Textgattungen ist alles andere als eindeutig – auch wenn sie im sozialen Gebrauch immer wieder der temporären Vereindeutigung bedarf. Wissenschaftler:innen, die sich mit alten und/oder fremdsprachigen Texten befassen, stehen auch forschungspraktisch immer wieder vor dem Problem, dass Texte und Textpassagen vieles bedeuten können und vermeintliche Eindeutigkeit erst durch den sogenannten »Kontext« hergestellt werden kann, also durch pragmatische kommunikative Einbettung, die wiederum oftmals mühsam rekonstruiert werden muss. Methodische Herangehensweisen wie die Sequenzanalyse zeigen darüber hinaus, dass auch einfache, auf den ersten Blick eindeutige Textpassagen aus bekannten Zusammenhängen prinzipiell polysem sind und nur durch ihre unmittelbaren pragmatischen Anschlüsse zeitweise vereindeutigt werden.

Kurz: Bilder und Texte sind gleichermaßen viel- und eindeutig. Kein Zeichen, ob bildlich oder textlich, ist per se eindeutig, sondern muss im konkreten empirischen Gebrauchszusammenhang temporär vereindeutigt werden. Manchmal wird die Vieldeutigkeit auch empirisch thematisiert und dann bleibt ein Zeichen »mehrdeutig«. Und auch die Mehrdeutigkeit von Texten wird – gerade religionsgeschichtlich – oftmals expliziter Teil der empirisch vorfindlichen Kommunikation über diese Texte. So gibt es seit Jahrhunderten innerhalb der verschiedenen religiösen Tradi-

tionen Auslegungsstreits über die rechte Deutung bestimmter textlicher Überlieferungen.

**Unmittelbare Affektion vs. mittelbares Verstehen** Es ist schon mehrfach angedeutet worden, dass in der bildwissenschaftlichen und daran orientierten Methodenliteratur Bilder als Gegenstände in der Welt aufgefasst werden, die »unmittelbarer« und suggestiver als Texte seien (z. B. Lengwiler 2011, 133–136). Auch die bereits zitierte Passage bei Banks formuliert diesen Aspekt (Bilder seien »messy«, »rich«, »particular« [2014, 405]). Bräunlein notiert mit Bezug auf Hans G. Kippenberg und Susanne Langer ebenfalls, dass nicht-sprachliche Sinneserfahrung autonom sei und Bilder daher primär »gefühlsmäßig« wirken (Bräunlein 2004, 210). Umgekehrt müssten also Texte kognitiv wirken und dabei sachlich-distanziert argumentieren.

Zum einen reproduziert diese These die längst überkommen geglaubten Dichotomien von Verstand/Gefühl, Geist/Körper, Kultur/Natur oder Rationalität/Irrationalität und setzt das Bild auf die Seite des Irrationalen und Suggestiven, zum anderen – und wichtiger für mein Argument – verkennt sie die empirische Vielfalt von Texten und Bildern. Es gibt Bilder, die dazu neigen, sachlich-rationale Informationen anzubieten (z. B. in naturwissenschaftlichen oder medizinischen Zusammenhängen), es gibt aber auch Texte, die dazu tendieren, suggestiv und affektiv zu wirken (z. B. politische Reden oder Lyrik). Texte sind somit wieder den Bildern viel näher als häufig angenommen: Sie können unmittelbar affektiv wirken und dabei suggestiv sein.

**Nicht-Sprachlichkeit vs. Sprachlichkeit** Es ist weitgehend *common sense* – sowohl in der Alltagswelt als auch in der bildwissenschaftlichen Literatur –, dass Bilder nicht-sprachlich sind (z. B. Lengwiler 2011, 133–136). Dabei bleibt häufig offen, was genau mit »Sprache« gemeint ist. Stattdessen argumentieren viele Autor:innen, dass der sogenannte *linguistic turn* in den Sozial- und Geisteswissenschaften mit dazu beigetragen habe, Bilder als Daten zu marginalisieren. Bohnsack vermutet, dass aufgrund eines »Paradigmas der Textinterpretation« die Untersuchung von Bildern »von vornherein weniger valide erscheinen« muss, weil Bilder, »wenn sie wissenschaftlich relevant werden sollen, grundsätzlich erst einmal in Beobachtungssätze beziehungsweise -texte umzuformulieren« sind (Bohnsack 2013, 63–64). Anders gesagt: Es ist zwar denkbar, in Bildern über Bilder zu kommunizieren, das System »Wissenschaft« verwendet aber in der Regel geschriebene Sprache (Texte), und daher ist die methodologisch geleitete Überführung von Bildern in Texte unvermeidlich (auch die bildwissenschaftliche Literatur füllt inzwischen Tausende von Seiten mit Text). Dass dadurch Bildanalysen weniger valide seien, scheint mir eine unbegründete Vermutung, die darüber hinaus die erwähnten Dichotomien von Vernunft/Gefühl, Geist/Körper, Kultur/Natur reproduziert. Eine Textanalyse, die ihre Daten, also geschriebene Worte und Sätze, methodisch unkontrolliert

interpretiert, ist genauso wenig valide wie eine methodisch unsaubere Bildanalyse. Die Herausforderung besteht meiner Ansicht nach darin, die Datengattung »Bild« gegenstandsadäquat in wissenschaftlicher Kommunikation zu bearbeiten.

Um die Differenz von Sprache und Bild zu erläutern, wird in der Literatur ausgeführt, dass Bilder beispielsweise kein Alphabet, kein Lexikon und keine Grammatik haben, keine »distinkte[n] Grundeinheiten« wie Buchstaben, Silben oder Wörter und somit auch keine »stabilen bedeutungstragenden und -unterscheidenden Einheiten«, die als Zeichen erkannt und verstanden werden könnten (Grave 2014, 63–64). »Was im Kontext einer bildlichen Darstellung als Motiv, Gegenstand oder Figur identifiziert wird«, so fährt Johannes Grave fort, »muss vielmehr zunächst aus dem Kontinuum der visuellen Phänomene im Bild hervortreten und durch den Betrachter abgegrenzt werden« (Grave 2014, 63–64). Das ist zwar richtig, gilt aber auch für geschriebene und gesprochene Sprache: Schon de Saussure hat gezeigt, dass ein (textliches) Zeichen erst dadurch ein Zeichen wird, dass es sich von anderen unterscheidet, und nicht, weil es *a priori* als Zeichen da ist: »Sprache [...] enthält nur begriffliche und lautliche *Verschiedenheiten*, die sich aus dem System ergeben. Was ein Zeichen an Vorstellung oder Lautmaterial enthält ist weniger wichtig als das, was *in Gestalt der andern Zeichen um dieses herum gelagert ist*« (Saussure 1967 [1916], 143–144; Hervorhebung MR). Mit anderen Worten: Es ist nicht selbstverständlich, dass eine Menge von schwarzen Linien auf Papier ein Wort darstellt, das mit einem bestimmten lautlichen Klang korrespondiert, der wiederum ein gewisses semantisches Potenzial anbietet. Was im Kontext einer schriftlichen (oder mündlichen) Darstellung als Wort identifiziert wird, muss zunächst aus dem Kontinuum der visuellen (oder auditiven) Elemente im Text hervortreten und durch die Betrachter:innen abgegrenzt werden – das weiß jede:r, der einmal lesen gelernt hat oder eine unbekannte Schrift entziffern will.

Aus der Vermutung, dass Bilder nicht sprachlich seien, folgt für viele Autor:innen der Schluss, dass semiotische Ansätze für Bilder nicht fruchtbar seien (z. B. Bräunlein 2009, 773). So fasst Tim Karis einen zentralen Aspekt der Visual Culture Studies wie folgt zusammen:

»Die Vorstellung von Bildern als Zeichen bzw. als Phänomenen, die aus einzelnen distinkten Zeichen aufgebaut sind, die jeweils benennbar wären, wird in der Perspektive der Visual Culture Studies als Relikt des frühen Strukturalismus verworfen und durch ein dynamisches Verständnis von Zeichensystemen ersetzt. Die Bedeutung eines Bildes ist folglich nicht *per se* gegeben, sondern ergibt sich aus dem zeitlichen und räumlich-kulturellen Kontext seines Erscheinens« (Karis 2013, 138).

Und vor dem Hintergrund es eben Erörterten müsste ergänzt werden: Das gleiche gilt auch für geschriebene und gesprochene Sprache: Die Bedeutung eines sprachlichen Zeichens oder einer Menge von sprachlichen Zeichen ist nicht *per se* gegeben,

sondern entsteht erst in einem Arrangement von Zeichen. Nicht umsonst geben Wörterbücher für ein fremdsprachiges Wort in der Regel mehrere deutschsprachige Entsprechungen an und überlassen es den Übersetzer:innen, die jeweils »passende« je nach »Kontext« (d.h. Arrangement) der zu übersetzenden Stelle auszuwählen

Karis resümiert weiter: »Bilder sind nicht [...], wie es die am Strukturalismus orientierte Semiotik glaubte, als arbiträre Zeichen in einem Bildsystem nur durch ihr Verhältnis zu anderen Zeichen zu erfassen. Vielmehr sind Bilder nur aus ihrem Gebrauch heraus zu bestimmen« (Karis 2013, 140; ähnlich auch Bräunlein 2009, 777). Man kann diesen vermeintlichen Widerspruch (Verhältnis zu anderen Zeichen in einem Bildsystem vs. Gebrauch von Bildern), so meine Vermutung, auflösen, da ein konkretes *Verhältnis von Zeichen zueinander* immer nur im *Gebrauch*, also in der Pragmatik, hergestellt wird.

Der Ansatz, dass Bilder nicht »semiotisch« seien, wendet sich auch gegen die sogenannte »Ähnlichkeitstheorie«, die – so formuliert beispielsweise Bräunlein (2009, 776) unter Rekurs auf Nelson Goodman (1976) – davon ausgeht, dass die Bedeutung von Bildern auf Ähnlichkeit mit dem Abgebildeten basiert. Dem liegt, so meine Ansicht, ein verengtes Verständnis von Semiotik, genauer: der Peirce'schen Semiotik, zugrunde. Peirce geht nicht davon aus, dass die Zeichenaspekte (indexikalisch, ikonisch, symbolisch) kategorisch auf verschiedene Zeichentypen anwendbar seien, dass es also »ikonische Zeichen« (wie z. B. Bilder) und »symbolische Zeichen« (wie z. B. Worte) gebe. Vielmehr weisen alle Zeichen diese Aspekte oder Eigenschaften auf (Krech 2012a, 52). Bilder funktionieren demnach zwar häufig ikonisch, weil sie unter anderem in einer Ähnlichkeitsbeziehung zum Abgebildeten (Objekt, Referent oder Signifikat) stehen, aber sie sind zugleich symbolisch, d.h. arbiträr und konventionell, und indexikalisch, weil sie auf das hinweisen, was sie abbilden. Eine Semiotik Peirce'scher Prägung mit systemtheoretischer Grundierung konzipiert nicht sämtliche Zeichentypen »sprachanalogue« (Bräunlein 2009, 772), sondern löst sich von der Sprache und geht damit der vermeintlichen Gefahr eines »Linguismus« (Bräunlein 2009, 773) aus dem Weg.

Die Rede von der Dominanz des *linguistic turn* verkennt meiner Ansicht nach darüber hinaus, was einen *turn* ausmacht: Der *linguistic turn* fordert nicht, nur noch Sprache zu untersuchen – genauso wie der *performative turn* nicht fordert, nur noch Performanzen zu untersuchen, oder wie der *spatial turn* nicht fordert, nur noch Räume zu untersuchen. Der *linguistic turn* fordert auch nicht, alle empirisch vorfindlichen Daten so zu untersuchen, *als ob* sie Sprache seien. Eine erkenntnisfördernde Neuausrichtung von Wissenschaftsgebieten fordert im Gegenteil, aus dem Gegenstandsbezug (z. B. auf Sprache, Räume, Rituale oder Bilder) eine theoretische Perspektive und ein konzeptionelles wie methodologisches Instrumentarium zu entwickeln: »So werden beispielsweise »Ritual«, »Übersetzung« oder »Raum« von *Forschungsgegenständen zu Analyse kategorien*, mit denen dann auch Phänomene erfasst werden können, die ursprünglich nicht in den traditionellen Gegenstandsbereich im enge-

ren Sinn gehören« (Bachmann-Medick 2007, 26; Hervorhebung MR). Ein so verstandener *semiotic turn* untersucht *nicht* nur noch Texte oder textliche Zeichen; er verlangt auch *nicht*, alle empirisch und historisch auffindbaren Gegenstände so zu untersuchen, *als ob* sie Texte oder Zeichen seien. Aber er fordert, die Erkenntnisgewinne aus den Sprachwissenschaften und aus der Semiotik auf sozio-kulturelle Phänomene insgesamt zu übertragen. Dann geht es um die Rekonstruktion und Erklärung von sozio-kulturellen Bedeutungsprozessen. Und hierfür sind Begriffe und Erklärungsansätze aus semiotischen Theorien, so die hier vorgeschlagenen Überlegungen, besonders hilfreich.

Die Kritik an einem semiotischen Zugang zu Bildern mag berechtigt sein, wenn man eine enge Auffassung von Semiotik zugrunde legt, der zufolge es in geschriebener und gesprochener Sprache identifizierbare Zeichen gibt, denen man eine genaue Bedeutung zuweisen kann. Der Bildwissenschaftler Martin Schulz (2005, 77) formuliert diesen vermeintlichen »Linguismus« semiotischer Zugänge zum Bild sehr deutlich:

»Gibt es überhaupt eine den abstrakten und stark codierten Zeichensystemen der Schrift vergleichbare Systematik des Bildes, eine syntaktische Struktur, die sich ähnlich in diskrete Einheiten zerlegen ließe? [...] Muss denn ein Bild immer ein sinnhaftes Zeichen sein, um als Bild erkannt werden zu können?«

Schulz stellt fest:

»[S]emiotische Analysen [scheinen] dahin zu tendieren [...] Bilder vor allem als gespeicherte, decodierbare, kontextabhängige Botschaften zu lesen, die einen bestimmten messbaren Wert in sich tragen, als wären sie gleichsam körperlose Zeichen für eine ebenso körperlos gedachte reine Kognition« (zitiert in Bräunlein 2009, 773).

Dem könnte man eine Reihe von ähnlichen Fragen erwidern, wenn man die hier formulierte prinzipielle Dichotomie von Bild und Text aufbrechen will: Was sind die diskreten Einheiten der geschriebenen Sprache? Gibt es so etwas wie kleinste bedeutungstragende Elemente in der gesprochenen und geschriebenen Sprache? Theoretisch gibt es sie in Gestalt der sogenannten Morpheme (z. B. Volli 2002, 59), im konkreten Gebrauchszusammenhang, beispielsweise in einem Satz, hilft die lexikalische Bedeutung eines solchen kleinsten bedeutungstragenden Elementes der Sprache aber oft nicht viel weiter (wie man aus der Übersetzungspraxis zur Genüge weiß).

Muss ein Text ein »sinnhaftes Zeichen« sein, um als Text erkannt zu werden? Nicht nur die Religionsgeschichte zeigt, dass auch Texte neben der sinnhaften Dimension (sie stehen im semantischen Sinne für etwas) auch eine ästhetische Di-

mension haben (sie stehen im ikonischen Sinne für sich selbst): Hier wäre zu denken an Kalligraphie, Initialen und Versalien in Manuskripten und von Hand kolorierten Drucken, an geschmückte und verzierte Bücher, die gezeigt, verehrt und als materielle Objekte bearbeitet werden. Das geht so weit, dass es Texte gibt, die nur »vorgeben«, Texte zu sein, zum Beispiel textähnliche Talismane des frühen Daoismus; das sind »eigenartige und bildhafte Kompositionen von chinesischen Schriftzeichen«, die als »Verschmelzung des Mediums der Schrift mit dem Medium des Bildes« zu verstehen sind, »ohne im engeren Sinne etwas bildlich darzustellen oder textlich aufzuschreiben« (Di Giacinto 2019, 59).

Schließlich ist mir keine semiotische Analyse bekannt, die etwas »messen« will – wie Schulz im obigen Zitat andeutet. In dieser Kritik am semiotischen Zugriff schlägt sich erneut die erwähnte Dichotomie von Verstand vs. Gefühl nieder: Texte und Textanalysen sind dann vermeintlich rational, während Bilder und Bildanalysen vermeintlich emotional sind. Eine Semiotik Peirce'scher Prägung kann nicht – wie von Schulz angedeutet – behaupten, Zeichen trügen »einen bestimmten messbaren Wert in sich [...], als wären sie gleichsam körperlose Zeichen für eine ebenso körperlos gedachte reine Kognition« (Schulz zitiert in Bräunlein 2009, 773). Die Semiotik weiß, dass Zeichen nicht »körperlos« sind: Jedes Zeichen braucht eine materielle Trägersubstanz (auch Zeichenmittel oder Repräsentamen genannt), seien es Papier und Tinte, der menschliche Stimm- und Hörapparat, oder Leinwand und Farbe. Die materielle Präsenz der Zeichen hat ohne Zweifel eine leiblich-affektive Wirkung auf wahrnehmende Menschen. Diese spielt sich aber zuerst im individuellen, situationsgebundenen Erleben ab und wird erst dann zu einem sozialen – und als solches: zu einem sozialwissenschaftlich untersuchbaren – Sachverhalt, wenn sie (verbal oder non-verbal, schriftlich oder nicht-schriftlich) kommuniziert wird, also sinnhafte Anschlüsse erzeugt. Ob dies zutrifft, ist nicht pauschal zu beantworten, sondern immer nur mit Blick auf den konkreten empirischen Sachverhalt.

Eng verbunden mit der Vorstellung, dass Bilder nicht-sprachlich sind, ist die Idee, dass es sich um einen Fall »präsentativer Symbolisierung« (nach Langer 1957) handelt und damit um »nicht-sprachliche Sinneserfahrung«: »Nicht begriffliche Symbolik, nicht sprachgeleitete Introspektion stehen am Anfang aller Erkenntnis, sondern das gefühlsmäßige Erleben eröffnet das Tor zur Außenwelt« (Bräunlein 2004, 210). Dieser Beobachtung ist aus alltagsweltlicher Perspektive zuzustimmen. Doch wird das »gefühlsmäßige Erleben« zum einen erst dann ein sozialer Sachverhalt und damit Gegenstand einer sozialwissenschaftlich orientierten Religionswissenschaft, wenn es – in welchem Medium auch immer – kommuniziert wird. Zum anderen ist auch jenes »gefühlsmäßige Erleben« immer schon sozial und kulturell bedingt und disponiert.

**Fazit** Was sich in den oben dargestellten Abgrenzungsversuchen bildwissenschaftlicher und verwandter Zugänge zur »Sprache« und zur »Semiotik« abzeichnet, ist ei-

ne implizite Aufrechterhaltung der Dichotomien von Geist vs. Körper, Vernunft vs. Gefühl, Kultur vs. Natur, Sprache vs. Bild sowie ein Missverständnis über den *linguistic* bzw. *semiotic turn*. Oft scheint der *linguistic turn* so verstanden zu werden, dass er Sprache (genauer: geschriebene Texte) als Gegenstand der sozial- und kulturwissenschaftlichen Forschung an erste Stelle setze. Der *semiotic turn* wird so rezipiert, als behauptete er, dass die soziale Wirklichkeit ›genau wie Sprache‹ funktioniere.

Doris Bachmann-Medick führt aus, dass ein *turn* zwar mit der ›(Wieder-)Entdeckung‹ eines Gegenstandsbereiches in der Empirie beginnt (z. B. Sprache, Ritual, Raum), der dann auch interdisziplinär Interesse weckt (Bachmann-Medick 2007, 26). Doch von einem *turn* im engeren Sinne könne man ›erst sprechen, wenn der neue Forschungsfokus von der Gegenstandsebene neuartiger Untersuchungsfelder auf die Ebene von Analysekatoren und Konzepten ›umschlägt‹, wenn er also nicht mehr nur neue Erkenntnisobjekte ausweist, sondern selbst zum Erkenntnismittel und -medium wird« (Bachmann-Medick 2007, 26). Insofern bedeutet ein *semiotic turn* nicht, sich nur mit zeichenhaften Gegenständen wie Sprache und Texten zu befassen, sondern Analysekatoren zu entwickeln, »mit denen dann auch Phänomene erfasst werden können, die ursprünglich nicht in den traditionellen Gegenstandsbereich im engeren Sinn gehören« (Bachmann-Medick 2007, 26). Diesen Ansatz vertrete ich auch für die Semiotik insgesamt, die sich somit von ihrem ursprünglichen Gegenstandsbereich, der Sprache, löst und auf andere Gegenstandsbereiche übertragen werden kann, ohne diese unreflektiert sprachanalog aufzufassen. Weil es dabei um Religion als ein soziales System geht, das im systemtheoretischen Sinne ausschließlich kommunikativ operiert, muss Kommunikation der Vorgang sein, der empirisch untersucht wird – jedoch, wie oben betont, mit einem Blick auf sämtliche Medien von Kommunikation, Bilder, Texte, Körper und Architekturen eingeschlossen.

Besonders pointiert formuliert David Morgan die Kritik an einer semiotisch grundierten Analyse von Bildern: Text und Bild träten zwar empirisch oft zusammen auf, seien aber nicht verlustfrei ineinander überführbar: »So the idea that images might be treated as visual texts, as signs of meaning that are properly textual in nature, is a presumption« (Morgan 2020, 1–2) – eine Annahme (oder, wie Morgan schreibt, »Anmaßung«), die, soweit mir bekannt ist, von keiner Semiotik vertreten wird. Morgan vertritt dagegen die Ansicht, dass man die mediale und performative Dimension von Sprache, aber auch von Bildern, nicht von ihrer Bedeutung trennen könne: »Entanglement means inextricability such that to change the configuration is to change the meaning« (Morgan 2020, 1–2). Dass Bedeutung auch auf Medien und deren Verwendung in konkreten sozialen Zusammenhängen angewiesen ist, ist wiederum ein Allgemeinplatz der Semiotik, die auf das Wechselspiel von Semantik, Syntax und Pragmatik hinweist (z. B. Krech 2021, 77), so dass Morgen hier – trotz seiner ausdrücklichen Ablehnung semiotischer Zugriffe – im weiteren Sinne selbst semiotisch argumentiert.

Ein letzter Punkt, der nach meinem Kenntnisstand in den Bildwissenschaften nicht immer ausführlich reflektiert wird, ist der Umstand, dass auch Bilder – genau wie Texte – keine homogene Datengattung sind. Bilder liegen in der historischen und gegenwärtigen Lebenswelt in einer Fülle von Formen und Varianten vor, häufig sind sie sogar unmittelbar mit textlichen Zeichen (Buchstaben, Wörtern) verbunden (z. B. Versalien, Lithographien, Werbeplakate), so dass schon das empirische Material in Vergangenheit und Gegenwart keine einfache Unterscheidung von Texten und Bildern nahelegt. Bilder können abstrakte Strichzeichnungen sein, Fotografien, Gemälde, Skizzen oder Piktogramme; sie können alleinstehen oder in Narrativen verbunden sein (z. B. *Biblia pauperum*, Comics, Filme). Letztlich macht dies die Diskussion um Texte vs. Bilder müßig, sind sie doch – wie auch andere Medien (z. B. Klang und Musik, Rituale) – Möglichkeiten der soziokulturellen Bedeutungsproduktion, die es allesamt gegenstandsadäquat zu untersuchen gilt.

Für die konkreten Untersuchungen und die Arbeit mit dem empirischen Material sind diese Ausführungen relevant, weil sie einen theoretischen und methodologischen Rahmen bieten, die vielfältigen Protokolle sozial-räumlicher Situationen (fotografische Protokolle, Texte, Interviewtranskripte) allesamt als ›konservierte‹ Kommunikation aufzufassen, die aus Elementen besteht, welche wechselseitig aufeinander verweisen.

