

---

# 9

## Contexts

---

The intent of this chapter is to provide capsule reviews of many of the main traditions of content analysis research. The summaries are intended to hit the highlights of common applications and major trends and to refer the reader to optimal resources for learning about the use of content analysis in a variety of contexts. The reader is reminded that certain other newer media contexts (e.g., video gaming) were summarized in Chapter 7.

The term *context* is used flexibly here. There are substantive contexts, such as the study of gender roles across communication forms—for example, in human interaction and in the media. There are medium-specific contexts, such as the emerging research on the nature of online social media (see Chapter 7) and the growing body of work on narrative films. And there are ideological contexts, such as the semantic-network approach to the study of linguistics. These are all ways in which people have framed content analysis and therefore categories in which bodies of research have accumulated. It should be noted that these contexts may overlap. That is, the context of minority images, for example, is composed of research in a variety of disciplines, looking at images in many different media (including film).

Generally, the contexts presented in this chapter are scholarly, but the chapter also includes a section on known applied and commercial applications of content analysis. Examples of these are harder to find than one might think, due to the proprietary nature of much of the commercial and government work. Some of the contexts of study require a specialized knowledge base for the researcher—linguistic analyses, psychological diagnoses, and musicological analyses of melody, for example. The references given should point the reader in the right direction for learning more about such specialized bodies of knowledge. For some of the contexts, good reviews of the literature to date are available and will be referenced. Taken together, the various sections of this chapter present a snapshot of the critical history of and the state of the art in content analysis.

## Psychometric Applications of Content Analysis

The goal of some content analysis work is to discover psychological characteristics about the individuals or groups who created the messages. A typical scenario is one in which a social researcher wants to make attributions about<sup>1</sup> psychological traits or states from written or spoken messages created by a series of individuals. Obviously, consideration of both psychology and linguistics is involved. It's no wonder that the two areas are often linked in the content analysis literature. When the goal is to measure these psychological characteristics, content analysis becomes a form of psychometrics (measures intended to tap internal, psychological constructs).

There are two main approaches to psychometric content analysis, both of which have been introduced in part elsewhere in this book: (a) thematic content analysis and (b) clinical content analysis. The two approaches overlap somewhat (e.g., both typically use measures of personality and mood factors), and they also share characteristics with studies from the literatures on open-ended coding, linguistic analyses, semantic networks, and stylometrics, as reviewed in the discussions to follow.

### Thematic Content Analysis

Thematic content analysis attempts to measure psychological characteristics of individuals, generally for survey or experimental purposes. The measures of the subjects' messages essentially stand in for other, alternative measures, such as self-report scales and indexes. The definitive early work on the use of computers to conduct this type of content analysis is Stone et al.'s (1966) edited volume *The General Inquirer: A Computer Approach to Content Analysis*. An essential source for this technique is *Motivation and Personality: Handbook of Thematic Content Analysis* (Smith, 1992), a fascinating and highly inclusive volume presenting no fewer than 80 different content analysis measurement schemes. Although the inclusiveness of the work means that some measures might be less well validated than others, the book is invaluable.

From Smith's (1992) perspective, thematic content analysis is the scoring of messages for content, style, or both for the purpose of assessing the characteristics or experiences of persons, groups, or historical periods (p. 1). This purpose is different than the goal of clinical work, where the outcome is a diagnosis of a psychological pathology. Here, the researcher is simply trying to measure person-based variables in an alternative fashion—without having the individuals fill out questionnaires. According to Smith, all the content analysis measures in the volume are oriented toward the use of previously validated individual-differences measures of “particular person variables” (p. 4). For example, the book includes such adapted measures as

Fear of success	(Fleming)
Leadership	(Veroff)
Helplessness	(Peterson)
The Protestant ethic	(McClelland & Koestner)
Job satisfaction	(McAdams)
Responsibility	(Winter)
Self-definition	(Stewart)

One well-used coding scheme for verbal samples (oral or written) is CAVE—content analysis of verbatim explanations (Peterson, Luborsky, & Seligman, 1983; Mehl, 2006). Its goal is to assess individuals' causal attributions, measuring whether they tend to find explanations for events that are internal versus external, stable versus unstable, and global versus specific (Peterson et al., 1988; Schulman et al., 1989). A related coding method is for pessimistic rumination, a measure of the frequency and intensity of pessimistic attributions, emotions, and events (Satterfield, 1998; Zullo, 1991).

We may also consider many of the dictionary measures of such CATA programs as LIWC and Profiler Plus to be thematic, psychometric content analysis applications (e.g., Tausczik & Pennebaker, 2010; see Resource 1). Further exploration has been conducted by Schwartz et al. (2013) in their analyses of a dataset of Facebook posts from 75,000 people who have also taken personality inventories; their results show the potential for identifying personality characteristics from word usage patterns in social media.

## Clinical Applications

Perhaps beginning with Freud's notion of "symptomatic texts" (Christie, 1999), psychologists and psychiatrists have been interested in tapping the inner workings of the troubled mind in an unobtrusive fashion, particularly by analyzing messages generated by that mind. With the diagnosis of psychological problems a goal, some researchers and clinicians have worked to develop appropriate content analytic techniques.

Lee and Peterson (1997) note that content analysis may be more than just an alternative to clinical diagnosis procedures, that in fact it may have advantages because it follows the scientific method:

What makes content analysis more than just a clinical impression from text is the explicit specification by the researcher of the rules used to make inferences. Content analysis requires that the researcher obtain suitable material, develop a coding protocol, and ascertain the

reliability and validity of actual coding. In this sense, content analysis is no different in principle from other research methods. (p. 960)

Probably the most published analyst with regard to the clinical application of psychological content analysis was Louis Gottschalk, who diagnosed Unabomber Ted Kaczynski as “not mentally impaired” based on the text of his manifesto (“Assessing Cognitive Impairment,” 1999) and President Reagan as experiencing a “significant increase” in cognitive impairment between the 1980 and 1984 presidential debates (although Gottschalk held the release of the findings until 1987; “Acclaimed Neuroscientist Pledges,” 1997). The books *Content Analysis of Verbal Behavior: New Findings and Clinical Applications* (Gottschalk, 1995) and *Computerized Content Analysis of Speech and Verbal Texts and Its Many Applications* (Gottschalk & Bechtel, 2008) are comprehensive sources of information on the use and validation of the 20 measurement schemes he and his colleagues developed first for human coding and, later, for computer coding via PCAD (see Resource 1). The Gottschalk-Gleser Scales include measures for anxiety, hostility, cognitive impairment, depression, hope, sociality, narcissism, achievement strivings, and positive affect.

## Open-Ended and Pictorial Responses

Content analysis is frequently applied to the task of organizing and assigning meaning to research participant responses to questionnaires, individual and group interviews, and researcher prompts. While one may choose to conduct qualitative analysis (Schreier, 2012), quantitative content analysis remains a popular choice. Both human and computer coding may be applied to the analysis of open-ended responses (Mohler & Zuell, 2001). In either instance, there are two types of approaches to coding individuals’ written and pictorial responses to questions and stimulus materials. One approach uses any of a variety of *preset coding* (a priori) schemes. When the goal is the measurement of psychological constructs, it essentially follows the thematic content analysis procedures described earlier (perhaps the only difference is that the measures in this section are typically applied to communicative responses to *specific* questions or tasks rather than analyses of fairly generic or naturally occurring speech or writing). The second approach uses *emergent coding* of the content, in which a coding scheme is established after all responses are collected; then, systematic content analysis is conducted applying this scheme to the responses, with appropriate reliability assessment.

Some of the earliest devised and still popular preset open-ended coding applications are for thematic apperceptive measures (the best known is the TAT, or thematic apperception test, of which there are many variations; Jenkins, 2008) and inkblot tests (such as the Rorschach). Thematic apperceptive measures require individuals to respond by making up stories to go along

with a standard set of pictures representing personal and interpersonal situations. Preset coding options include measures of need achievement, need affiliation, and need power (Chusmir, 1985). Inkblot tests ask the subjects to report what they see in each of a series of symmetrical, abstract inkblots. The responses may be scored with standard classifications for location (which portions of the blots are used), determinants (the form, or structural, characteristics commented on), and content (the substance of what is seen, such as animal or human referents). Additional coding options have been proposed, such as Urist's application of a mutuality-of-autonomy measure (measuring perceived psychological autonomy of others; Monroe et al., 2013; Urist, 1977).

Both types of measures were originally devised as projective tools for psychoanalysis, and some critics maintain that they cannot meet the standards of psychometric measurement; rather, they maintain, the techniques are idiographic and provide stimuli that will drive the clinical psychoanalytical process between analyst and subject (Aronow, Reznikoff, & Moreland, 1994; Gregory, 1987; Te'eni, 1998).

Focusing on *drawings* rather than written or transcribed responses to questions, the Human Figure Drawing literature in psychology seeks to use evaluations of drawings as psychological diagnostic tests (Koppitz, 1984). Similarly, the Goodenough-Harris Drawing Test is a psychographic technique for assessing the intellectual maturity of young people (Harris & Pinder, 1974). Others have used preset coding schemes to categorize drawings. Stiles, Gibbons, and Schnellmann (1987) used a standard 10-variable coding scheme to establish significant gender differences between ninth graders' drawings of the "ideal man" and the "ideal woman." DiCarlo et al. (2000), studying street children in Honduras, found both age and gender differences for an adapted "ideal person" coding scheme (see also Gibbons & Stiles, 2004). And Ames, Andsager et al. (2005) used content analysis to evaluate multiple aspects of respondents' drawn output for the draw-an-event test for risky sexual situations (DET-RS).

The second approach to open-ended responses, that of emergent coding, obviously uses coding schemes that are more idiosyncratic and less well validated but in many instances may be the required technique. When no useful standard classification or coding scheme exists, or when the researcher wishes to begin the development of a new scheme, the emergent option is employed. Crawford and Gressley (1991) faced just such a situation in their study of humor preferences and practices for men and women. Their study participants were asked to write a paragraph about a specific individual with an "outstanding sense of humor" (p. 222). Respondents' answers were transcribed and examined by the researchers, who discerned five theme dimensions: hostility, jokes, real-life humor, creativity, and caring. These five were used as separate variables, and each original written narrative was coded for the presence or absence of each. A high level of intercoder agreement was reached.

Ahmad et al. (2009) also derived their coding scheme from the communicative content at hand—physician–patient interaction in audiorecorded medical

visits. They derived a coding scheme focused on the detection of “intimate partner violence and control,” which they then applied via human coding. Fujioka (2005) used human coding to measure recall after experimental exposure to TV news by Mexican American and White American students. Knobloch (2008) conducted a survey on relational uncertainty and from open-ended responses derived 12 content areas—uncertainty about children, careers, finances, sex, retirement, religion, extended family, communication, leisure time, health, commitment, and household chores. And Chipperfield et al. (2009), in their study consisting of interviews with 353 adults aged 72 and older, performed an iterative process to derive major thematic categories that served as antecedents for nine selected emotions, and the reliability of their application was assessed with an independent coder.

## Linguistics and Semantic Networks

*Linguistics* is the study of the structure and nature of human speech. There is a range of emphases, from the structural, “formal” study of syntax to the more semantically focused study of language meaning (Markel, 1998). The latter overlaps with other areas of message study, particularly thematic content analysis for psychographic purposes.

Quantitative content analysis is only one of many approaches used in linguistics. Many cross-language comparisons and searches for universal characteristics of languages (Goldberg, 1981) use qualitative participant observation, ethnography, critical methodologies, or both. However, with computer advances, there has been a growing interest in quantitative linguistics (Kohler & Rieger, 1993), particularly in computational linguistics (Litkowski, 1992, 1999), and the general study of semantic grammars (systems that codify statements’ unique meanings-as-intended as opposed to surface-level, superficial, ambiguous syntax grammars; Roberts, 1997a).

Although technically a part of linguistics, the study of messages using semantic networks has its own distinct literature and adherents. An ultimate goal of this approach is to map the network of interrelationships among concepts. Carley (1993) has contrasted the typical use of text content analysis (focusing on the extraction of concepts from texts) with the use of semantic networks (or map analysis, focusing on both concepts and the relationships among them). Dictionaries may (or may not) be used for semantic network content analysis; the goal is to discover patterns of co-occurrence of words or phrases in a text via computer analyses (Young, 1996). Often, these patterns are displayed via 2-D or 3-D maps (e.g., using such programs as CATPAC), in a method similar to that of multidimensional scaling (see Figure 8.7 for an example; Barnett & Woelfel, 1988; Woelfel & Fink, 1980).

Like neural networks that represent cognitive structures and processes, semantic networks represent systems of concepts with “meaning-full”



relationships (Litkowski, 1999). These relationships are based on word usage by individuals, groups, organizations, or societies (Carley, 1997a; Doerfel & Barnett, 1999). As Carley (1997b) notes, “Language can be represented as a network of concepts and the relationships among them. This network can be thought of as the social structure of language or, equivalently, the representation of extant social knowledge” (p. 79). More and more, these semantic networks have been analyzed as being in association with cognitive meaning and/or social structure in applied settings (e.g., Diesner & Carley, 2005a, 2005b).<sup>2</sup>

## Stylometrics and Computer Literary Analysis

*Stylometrics* refers to the analysis of the *style* of language. The technique is used to identify a language style that is distinctive, with the aim of describing, distinguishing, and sometimes establishing authorship. Tweedie, Singh, and Holmes (1996) define style as “a set of measurable patterns which may be unique to an author” (p. 401), including the identification of the number of nouns or other parts of speech used, the number of unique words, and the most common words. Early authorship studies, dating back more than two millennia, used human-coded analyses, but as with semantic mapping, more sophisticated analyses have been made possible by advances in computer text analysis software and the ready availability of texts in electronic form (see Kucukyilmaz et al., 2008, for a review of contemporary statistical options).

Stylometrics was originally most commonly used in literary studies. One popular application has been to attempt to establish the authorship of texts or settle disputes about authorship. Shakespeare has received much attention, due to questions surrounding the true authorship of works bearing his name. Elliott and Valenza (1996) helped to develop more than 50 computer tests to compare writings attributed to Shakespeare to the writings of Shakespeare claimants—“nobler” authors who could have conceivably been the real Shakespeare. They compared textual features, such as preferred words, rare words, new words, grade level, contractions, intensifiers, prefixes, and suffixes, and concluded that *none* of the claimants matched Shakespeare. Another related application of stylometrics has been to *compare* the styles of authors. Sigelman and Jacoby (1996), for example, used computer analysis techniques to distinguish the works of mystery writer Raymond Chandler from his imitators, based on four main stylistic elements: simplicity, action, dialogue, and vivid language. They found Chandler’s work to be consistent in style and concluded that the imitators failed to successfully replicate that style. Potter (1991) offers a fairly comprehensive review and critique of stylometric and other content analyses of literature appearing in the journal *Computers and the Humanities* from 1966 through 1990.

In addition to its applications to literature, stylometrics has also been used in other humanities contexts. Whissell (1996) employed traditional stylometric procedures along with novel measures of emotionality to compare the songs of Beatles members Paul McCartney and John Lennon, discovering that Lennon was the less pleasant and sadder lyricist of the two. In a bizarre but intriguing application of stylometrics, Bucklow (1998) analyzed craquelure (the pattern of cracks that forms on paintings as they age) as a means of establishing authorship of works of art.

More recent work has used stylometric methods to attempt to identify anonymous online writers for various reasons. Kucukyilmaz et al. (2008) were able to correctly identify authorship 99.7% of the time in their study of peer-to-peer chat content. Abbasi, Chen, and Nunamaker (2008) successfully applied stylometrics and authorship identification to the posts of e-commerce traders, who may readily change their online identities to help manipulate their reputation (e.g., to sidestep “bad” reviews on eBay). And Harpalani et al. (2011) found an effective stylometric method to identify “unique language styles in vandalism” in order to better detect vandalizing activity on Wikipedia.

In sum, computers give literary, linguistic, and other scholars the opportunity to shed fresh light on the formal and stylistic features of important works and on the styles of average online communicators as well. With the continued growth and development of archives and computer text analysis sources, the future possibilities for the content analysis of literature and other works should be extraordinary.

---

## Interaction Analysis

---

The systematic analysis of human verbal interactions is a small but important part of the content analysis literature. Most studies of interactive discourse are *not* content analyses but rather more qualitative and ideographic in nature. However, a number of quantitative coding schemes have been developed. Notably, these schemes are typically called *interaction analysis* in the literature rather than *content analysis*.

In the communication field, perhaps the most prominent system is Rogers and Farace’s (1975) relational coding scheme. Owing a debt to the theoretic notions of anthropologist Gregory Bateson (1958) and to earlier classification systems, such as those by Bales (1950), Borke (1967, 1969), and Mark (1971), the Rogers and Farace (1975) scheme takes a relational communication perspective: It is concerned with the *control* or dominance aspects of message exchange in dyads. Each utterance in a verbal exchange is the unit of data collection. The scheme assumes that relational control is based on both the grammatical form of the message (the codes are assertion, question, talk-over, noncomplete, and other) and the metacommunicative response of the message relative to the statement that came before it



(the codes are support, nonsupport, extension, answer, order, disconfirmation, topic change, initiation-termination, and other). These codes are used to create control codes (one-up, one-down, and one-across), indicating movement by a speaker with regard to relational control.

The Rogers and Farace (1975) scheme has been used to analyze interactions between husbands and wives (Courtright, Millar, & Rogers-Millar, 1979; Rogers-Millar & Millar, 1978), employees and employers (Fairhurst et al., 1987), physicians and patients (Cecil, 1998; O'Hair, 1989), and parents and children (Seklemian, as cited in Cecil, 1998). An adapted system has even been applied to television character interactions (Greenberg & Neuendorf, 1980; Neuendorf & Abelman, 1987).<sup>3</sup>

The validity of the Rogers and Farace (1975) scheme has been questioned (Fairhurst & Cooren, 2004; Folger & Poole, 1982), and alternative schemes for the analysis of verbal interactions have been presented. They include Ellis's (1979) RELCOM system and Tracey and Ray's (1984) topic-based coding scheme and Patterson's (1982) Family Interaction Coding Scheme, designed specifically to tap aversive interactions in families with an antisocial child. The SYMLOG (SYstem for the Multiple Level Observation of Groups) system for analyzing group interactions has many elements that are similar to quantitative content analysis, but ultimately, the measures rely on observers' subjective evaluations rather than objective, reliable estimates (Bales & Cohen, 1979).

MacWhinney's (1996; 2000) CHILDES project (the Child Language Data Exchange System), dedicated to the study of language learning, includes CHAT, a system for discourse notation and coding. This scheme calls for the separate coding of each utterance for phonology, speech acts, speech errors, morphology, and syntax. The widely used Roter Interaction Analysis System (Beach et al., 2011; Roter, Hall, & Aoki, 2002; Roter, Lipkin, & Dorsgaard, 1991; Vail et al., 2011), designed for the coding of interactions between doctors and patients, codes each statement or complete thought into one of 34 categories (e.g., agreement, worry, procedural instructions, persuasive attempts regarding therapeutic regimen). The Marital Interaction Coding System (Heyman, Weiss, & Eddy, 1995) and the Dyadic Parent-Child Interaction Coding System (Eyberg & Robinson, 1983) are both designed to analyze family interactions for family therapy purposes. The Brief Romantic Relationship Interaction Coding Scheme (BRRICS) is designed to assess aspects of a romantic relationship "quickly and efficiently" (Humbad et al., 2011).

A number of other interaction analysis schemes have been proposed (e.g., Gunawardena, Lowe, & Anderson, 1997; Hirokawa, 1988; Jones et al., 1999; Tardy, 1988; Zemel, Xhafa, & Cakir, 2007). Generally, the coding schemes have been applied to specific domains of interaction, such as group decision-making, negotiation, family communication, communication in romantic relationships, or classroom interaction, with measures designed to provide in-depth understanding of the nuances of communication within a given context.<sup>4</sup>

## Other Interpersonal Behaviors

Similarly, a number of rigorous and sound coding schemes have been developed for the systematic analysis of other interpersonal communication behaviors, not generally applied to a representative sample of realistic interactions but used for measuring actions in experimental settings. In addition, the coding of human behaviors in the study of psychological processes might also be deemed content analytic, and that literature can inform us with regard to unitizing and intercoder reliability (Bakeman, 2000).

The study of nonverbal behaviors has produced primarily qualitative, idiographic, and gestalt-type investigations. However, taxonomies (category schemes) for the study of nonverbals are quite well developed (Ekman & Rosenberg, 1997; Messing & Campbell, 1999). And some researchers have used systematic and objective coding schemes looking at paralanguage (e.g., speech rate: Beaumont, 1995; Feyereisen & Harvard, 1999; Kelly & Conture, 1992), body positioning (e.g., Newton, Engquist, & Bois, 1977), facial expressions (Lagerspetz, Wahlroos, & Wendelin, 1978), and gestures (e.g., Feyereisen & Harvard, 1999; Guerrero & Burgoon, 1996; see also the *Journal of Nonverbal Behavior* and the general text on nonverbal communication by Knapp, Hall, & Horgan, 2014). The Facial Action Coding System (FACS) has been developed for the identification of the physical, facial expression of emotions (Ekman, Friesen, & Hager, 2002); a computer automation of the FACS scheme has been developed (Hamm et al., 2011).

A number of studies have looked at nonverbal behaviors for such mediated persons as characters in commercials aimed at children (Browne, 1998), candidates in political commercials (Hacker & Swan, 1992), televangelists (Neuendorf & Abelman, 1986), and the film roles of a durable screen actress (i.e., Bette Davis; Ealy, 1991). Others have examined the nonverbal displays in profile photos of private citizens on social media, including eye gaze, head angle, mouth position, general posture, and hand/finger placement (Hum et al., 2011; Kane et al., 2009).

Self-disclosure, the sharing of personal information with another, has been studied, with age and gender differences a main focus. Although the evidence on age had been mixed (Capwell, 1997; Collins & Gould, 1994), the evidence with regard to gender seems clear—women are more likely to disclose than are men, especially when discussing intimate topics (Cozby, 1973; Dominick, 1999; Hacker, 1981; Shaffer, Pegalis, & Cornell, 1991; Trammell & Keshelashvili, 2005).

Interrupting behaviors have been studied by a number of content analysts. Researchers have found conversation interruptions to occur more often when one is speaking to someone of the opposite sex rather than someone of one's own sex (Dindia, 1987) and to be more frequently committed by adolescent girls than their mothers (Beaumont, 1995). The long-held belief that males interrupt more than do females has been supported by

some studies (e.g., Zimmerman & West, 1975; Hancock & Rubin, 2015) and refuted by others (e.g., Dindia, 1987; Marche & Peterson, 1993).

---

## Computer-Mediated Communication

Some scholars contend that the online environment provides a particularly valid and neutral locus for unfettered, natural communication between individuals. Following early fears of a limited “cues filtered out” model of communication via computer (i.e., whereby most nonverbal cues such as tone of voice, use of gestures, and proxemics are lost in computer-mediated interaction; Sproull & Kiesler, 1986; Walther & Parks, 2002), scholars began to recognize the ways in which the “so-called impediments of communication technology are overcome by its users” (Walther, 2004, p. 386). The “hyperpersonal” model stands in contrast to the limited model, suggesting that computer-mediated communication (CMC) may actually facilitate social interaction because communicators may take more time and greater care in creating messages than they would face-to-face (FtF; Duthler, 2006; Walther, 2007).

Newton, Kramer, and McIntosh’s (2009) CATA (i.e., LIWC) study of blogging by individuals with autism spectrum disorders (ASP) found similarities in language choice between ASP and neurotypical bloggers that suggest the communication deficits exhibited by ASP communicators in face-to-face contexts are due to social contextual cues. The asynchronous nature of CMC may offer the user a “place of refuge” for communicative expression, and, as noted by Kim et al. (2007), online communication may facilitate communication for individuals who are shy or might otherwise be marginalized in FtF interactions. Plus, asynchronous communication provides additional overtime factors for potential analyses (e.g., Angeli & Schwartz, 2016, in press; De Wever et al., 2010; Zhou et al., 2004).

New modes of communication may engender new linguistic and communication norms. CMC research reveals striking differences in how online text differs from non-CMC writing, as summarized by Abbasi and Chen (2008): CMC is richer in interaction (for both synchronous and asynchronous forms), CMC is less topical, and CMC technologies allow the emergence of novel language varieties (p. 813). Recognizing the evolving nature of online language, Neylarouskaya, Prendinger, and Ishizuka (2007, 2009) began to supplement natural language processing techniques with the inclusion of web-centric “symbolic cues” such as emoticons and abbreviations in their analyses of source affect/emotion/sentiment as expressed through online writings.

---

## Organizational Communication

Content analysis has been applied in the organizational context with growing frequency over the years; content analyses have studied communication

patterns within the organization as well as communications of the organization with outside entities and publics. For internal communications, content analysis procedures have historically been a vital part of organizational communication audits (e.g., Downs & Adrian, 2004; Greenbaum, 1974), which also include surveys, interviews, and focus groups. An example of another application of content analysis to internal organizational communication is Downing's (2007) analysis of American Airlines' corporate strategies and tactics when communicating with its employees following the 9/11 attack.

An early set of literature studied the substance of e-mail, voice mail, and other organizational technologies used within and outside the organization. Just a few studies examined the content of voice mail (e.g., Rice & Danowski, 1991) and computer conferencing (e.g., Rosenbaum & Snyder, 1991). A greater number of studies looked at e-mail content generated in educational and business contexts (Danowski & Edison-Swift, 1985; Hill, Camden, & Clair, 1988; Kot, 1999; Marttunen, 1997; McCormick & McCormick, 1992). Generally, these e-mail studies identified social support communication networks that linked together individuals based on interests rather than physical location, sometimes creating links and relationships that did not exist before the introduction of e-mail. They also identified communication strategies used in e-mail that are different than those used in face-to-face interaction (e.g., Pratt et al., 1999). And both the content of e-mails and the patterns of contact have been the subject of examination, as in Debreceeny and Gray (2011), which demonstrated the utility of the data mining of e-mails as part of an accounting audit in order to perform forensic tasks such as attributing authorship/identity of the writer, evidentiary discovery in legal matters, fraud detection, and social network analysis among organizational teams and individuals.

Content analysis has also been used often to study annual reports, mission statements, corporate responsibility statements, and other organization-to-public communicative venues. For example, Holder-Webb et al. (2008) used content analysis to first identify the reporting formats for the release of mandatory and government-recommended corporate governance information to the public and then to analyze types of information disclosed as related to firm size and board type among a sample of 50 U.S. corporations. Hooghiemstra (2008) compared annual reports cross-culturally, concluding that both Japanese and U.S. corporations "explain company results in a self-serving fashion" (p. 618). Chizema (2008) identified significant predictors of German corporations' disclosure of pay rate for individual executives in annual reports. Short and Palmer (2008) used the CATA program Diction in concert with human coding to analyze mission statements from 408 schools of business, finding significant differences in word usage based on such factors as the presence of key mission statement elements and the school's national ranking. Campopiano and De Massis (2015) examined how family influence on business organizations relates to corporate social responsibility reporting, analyzing the CSR content of 98 Italian firms. They found that in comparison to nonfamily firms, family firms disseminate a greater variety of CSR reports, are

less compliant with accepted CSR standards, and tend to emphasize different CSR topics (with family firms placing less emphasis on “values and general interests,” “shareholders,” “employees,” and “customers,” and more emphasis on “environmental and green issues,” and “philanthropy”). Stohl, Stohl, and Popova (2009) examined the degree to which the codes of ethics of 157 world corporations on the Global 500 and/or *Fortune* 500 lists included elements of the “third generation” of corporate social responsibility, finding some evidence of global convergence but also a confirmation of corporations’ primary concern with profits and government-mandated organizational behaviors.

Of course, the use of newer media by organizations has also been studied—including the use of Twitter by disaster relief organizations following the 2010 Haiti earthquake (Gurman & Ellenberger, 2015), the use of YouTube by nonprofit organizations to inform and educate the public (Waters & Jones, 2011), and the degree of incorporation of the Japanese “5S” philosophy of life into Japanese, U.K., and U.S. corporate web sites (Kobayashi, Fisher, & Gapp, 2008).

---

## Health Communication

Content analysis has been used in a variety of ways in the health arena (Jordan et al., 2009; Tian & Robinson, 2014). There have been content analyses of physician interactions, communication within health communities, public health campaigns, and, most prominently, of health images in the media.

The communicative behavior of physicians in real life has sometimes been content analyzed. For example, La Pean and Farrell (2005) analyzed transcribed statements made by pediatric residents to parents of infants found through newborn genetic testing to be carriers for sickle cell hemoglobinopathy (SCH) or cystic fibrosis (CF). They found 69% of the statements to be “initially misleading”—that is, to deliver “bad news” elements prior to the delivery of the “good news” of the child’s full health. Eggly et al. (2006) examined questions posed by patients and their companions during outpatient interactions with oncologists who had to deliver “bad news,” with companions asking more questions than patients. And Eggly, Brennan, and Wiese-Rometsch (2005) content analyzed self-report essays by medical interns regarding professional and unprofessional behaviors they had encountered during their medical education, confirming a prevalence of professionalism.

The communication patterns of health communities have also been content analyzed. Buis (2008) studied posts in an online hospice support community over a 4-month period, finding that emotional support interactions were far more frequent than informational support exchanges. Van Uder-Kraan et al. (2008) explored who uses Dutch online health support groups for breast cancer, fibromyalgia, and arthritis and what is exchanged by the participants, finding generally functional usage with few of the anticipated negative manifestations.

Health messages in the mass media have been subject to numerous content analyses. Reviews of this sizable body of research can be found in Neuendorf (1990b) and Manganello and Blake (2010); some exemplars follow. In their compilation of 441 content analyses of health messages in the media between 1985 and 2005, Manganello and Blake found the seven most prominent health topics subjected to content analysis across all media to be violence, sex, tobacco, obesity/body image, alcohol, cancer, and aging.

In studies of nonfiction media content, some analyses focus on tracking whether health news is reported accurately, as with studies of news coverage of the HPV vaccine (Habel, Liddon, & Stryker, 2009; Johnson, Stonean, & Scott, 2011), or what “tone” is prevalent in the treatment of medical information, as with a study of news coverage of fMRI (functional magnetic resonance imaging) technology when it was new (Racine, Bar-Ilan, & Illes, 2006). One unique study even tracked the over-time prevalence of themes of taste, tradition, and science (and therefore, nutrition and health) in introductory materials to the four different editions of the *Joy of Cooking* (Heiss & Bates, 2014), finding a dominance of science in all editions except the first in 1964.

Other analyses focus on representations of individuals in the health context, as in Gilbert, MacCauley, and Smale’s (1997) study of news representations of persons with disabilities. And Puhl et al. (2013) found online news videos to portray obese individuals primarily in a negative, stigmatizing manner, often showing them as “headless” and photographed from the rear. In an ambitious longitudinal study of Dutch nonfiction medical TV series produced between 1961 and 2000, Verhoeven (2008) found that the speaking time allotted to experts decreased over time, while that allotted to laypersons increased, with fewer references to sources and science and more expressions of emotion and tension. Therefore, Verhoeven identifies three eras of Dutch medical television: the scientific, journalistic, and lay periods.

The study of health-related advertising has included many content analyses. For example, drug advertising in medical journals has, over decades, continued to be sex-typed in its appeals, according to content analyses (e.g., Curry & O’Brien, 2006; Michelson, 1996). Studies of alcohol advertising have been abundant over the years (although more abundant in the 1970s and 1980s than currently; e.g., Horner, Jamieson, & Romer, 2008; Finn & Strickland, 1982; Morgenstern et al., 2015).

Content analyses of public health campaigns have looked at such things as how community-based organizations use social media channels (Facebook, Twitter, YouTube) for health promotions activities (Ramanadhan et al., 2013) and what cultural differences exist between American and South Korean weight loss web sites (Baek & Yu, 2009).

Fictional media treatments of health topics have also come under the scrutiny of content analysts. An impressive initiative by the Kaiser Foundation examined the health-related sexual behaviors of TV characters over a 6-year period (Kunkel et al., 2005). Other research has focused on such diverse topics as images of mental illness in soap operas (e.g., Fruth & Padderud, 1985),



portrayals of physical disability on children's television programming (Bond, 2013); alcohol consumption in cartoons between 1930 and 1996 (Klein & Shiffman, 2013); tobacco portrayals in advertising and entertainment media content (Dewhirst, 2008); the incidence of substance use and abuse in fictional entertainment media (Greenberg et al., 1980; Roberts & Christenson, 2000); images of elderly characters in children's books (Danowski, 2011); and portrayals of health professionals on TV (e.g., Gerbner, Morgan, & Signorielli, 1982).

---

## Advertising

There are two divergent approaches to the content analysis of advertising and similar promotional messages—the marketing–advertising professional approach and the social effects approach (see Kim et al., 2014, and Taylor, 2005). The professional approach seeks to understand the content and form of advertising in order to produce more effective promotional materials (e.g., Connaughton & Jarvis, 2004; Gagnard & Morris, 1988; James & VandenBergh, 1990; McQuarrie & Phillips, 2008; Naccarato & Neuendorf, 1998; Stewart & Furse, 1986; also see Box 2.2).

The social effects approach is motivated by concern over the impact of advertising on individuals and societies. A number of prominent areas of study have emerged. The role of women in advertising has received quite a bit of attention across time (e.g., Ferrante, Haynes, & Kingsley, 1988; Riffe, Place, & Mayo, 1993; Signorielli, McLeod, & Healy, 1994; Smith, 1994; Verhellen, Dens, & de Pelsmacker, 2016, in press). International comparisons of advertising styles and substance have been an important part of the literature (e.g., Cheng & Schweitzer, 1996; Kalliny et al., 2008; Murray & Murray, 1996; Prominski, 2006; Tak, Kaid, & Lee, 1997; Wongthongsri, 1993). The tracking of stereotypes of minorities and the elderly in advertising (e.g., Bailey, 2006; Mastro & Stern, 2003; Miller, Leyell, & Mazachek, 2004; Prieler et al., 2015; Taylor & Bang, 1997), the analysis of advertising for potentially hazardous products, such as tobacco and alcohol (e.g., Belstock et al., 2008; Finn & Strickland, 1982; Lee & Callcott, 1994; Neuendorf, 1985, 1990a), and the critical examination of political advertising, notably negative strategies (e.g., Druckman, Kifer, & Parkin, 2010; Tak, Kaid, & Lee, 1997), have all continued to be viable areas of study. The validity of claims made in promotional messages has also been examined, as in Grana and Ling's (2014) study of e-cigarette commercial web sites.

A special focus has been on advertising aimed at children, almost always from the standpoint of the social effects scholar (e.g., Smith, 1994). For example, Rajewski et al. (1994) found a number of attractive but dubious themes in TV ads for food for kids—64% of the sample ads included some combination of violence, conflict, and trickery. Rose, Merchant, and Bakir (2012) found 83% of television ads for food products aimed at children employed

fantasy appeals, while qualitative interviews found children to be mixed in their ability to understand the manipulative intent of advertising.

## News

A tremendous number of studies have examined news content, nearly all using text analysis, either human coded or, much more often today, some type of CATA. The range of topics studied is large, overlapping with other context headings in this chapter, such as the treatment of minorities and women, and political communication.

The coverage of science issues has been a meaningful focus (e.g., Dudo, Dunwoody, & Scheufele, 2011; Dunwoody & Peters, 1992), with an eye to establishing whether scientific findings are reported fully and accurately. There are scores of studies tracking international news content and flow (see Shoemaker and Cohen's (2006) comprehensive 10-nation study). Past studies generally found that developing nations were well supplied with information about Western nations, whereas Western nations found comparatively little in their news about the rest of the world (Chang, 1998; Stevenson, 1994). But more recent studies have found the situation to be increasing complex (e.g., Wilke, Heimprecht, & Cohen, 2012), especially considering online news content. Some content analyses have taken a historical approach—in a longitudinal study of general news and social trends in the United States, Danielson and Lasorsa (1997) studied 100 years of front-page content in *The New York Times* and the *Los Angeles Times*, identifying such shifts as a decrease in emphasis on the individual and an increase in emphasis on the group and a move away from religion and local governmental power to expert authority and central government (p. 114).

Content analysis is often the method of choice for the numerous studies of news framing (Scheufele & Tewksbury, 2007; Shah et al., 2009). A news frame is a way in which a concept is communicated about in such a way as to encourage certain receiver interpretations and to discourage others. Looking at news sources, framing studies attempt to describe how news is presented by particular news outlets (Entman, 1993). From a receiver perspective, framing studies are designed to probe how members of the public may derive meaning from these presentations and how their opinions might be influenced (Semetko & Valkenburg, 2000). For example, Andsager and Powers (1999) found that *Newsweek* framed breast cancer stories with regard to causes and treatments, whereas *Time* more often used an economic framing (e.g., insurance concepts), and *U.S. News and World Report* presented breast cancer news with a research focus.

Other framing studies that have used content analysis include Grimm and Andsager's (2011) study of H.R. 4437, the 2005 proposed law restricting the rights of undocumented immigrants; Messner and South's (2011) study of major newspapers that found Wikipedia to be framed as credible

and accurate; Cissel's (2012) comparison of mainstream and alternative news framings of Occupy Wall Street; Colson's (2010) framing study of TV news coverage of the causes and prevalence of autism; Schwalbe, Silcock, and Keith's (2008) study of the visual framing of the U.S.-led invasion of Iraq in 2003, which found news images to shift from conflict to human interest; and Bennett, Lawrence, and Livingston's (2006) indexing of coverage of the Abu Ghraib scandal and Porpora, Nikolaev, and Hagemann's (2010) rebuttal to that framing analysis. Framing analyses have sometimes clustered or networked concepts based on co-occurrence, allowing interpretation as to what frames emerge from the mapped outcomes (e.g., Miller, Andsager, & Riechert, 1998; Miller & Riechert, 2001). For example, Qin (2015, p. 166) mapped coverage of Edward Snowden, and found him to be a "hero on Twitter" but a "traitor on [legacy] news."

---

## Political Communication

Researchers in the political arena made many early contributions to the methodology of content analysis (see "Milestones in Content Analysis History" at *The Content Analysis Guide Online* (CAGO)). This includes Harold Lasswell's prodigious contributions to quantitative approaches in the social and behavioral sciences, beginning with his 1927 dissertation, a "content analysis" of World War I propaganda techniques (see Lasswell, Leites, & Associates, 1949). Neuendorf and Kumar (2016, in press) have reviewed the use of content analysis in political communication, identifying two major strains of study—(a) works examining the political news function and (b) works looking at the campaign function (including such persuasive communications as politicians' speeches, political platforms, debates, political advertising, web sites, and social media; see Xenos and Foot (2005) for a comparison of online and offline political campaigning).

News coverage of political issues has often been studied from the agenda-setting viewpoint (McCombs, 2005). This theoretic perspective proposes that media content doesn't so much tell the audience *what* to think as it tells the audience *what* to think *about*. The news sets the public opinion agenda, bringing some issues to the forefront and minimizing others. This theory has motivated researchers to combine content analyses with public opinion survey data (e.g., Aalberg et al., 2013; Bayulgen & Arbatli, 2013; Fan, 1988; 1997; Hertog & Fan, 1995; McCombs et al., 1997; Pfau et al., 1998). Some studies have also examined the so-called intermedia agenda setting function, in which news coverage by one medium or media institution is followed by similar coverage by another medium or institution at a later date (Golan, 2006; Lim, 2011; Lopez-Escobar et al., 1998; Roberts & McCombs, 1994). Generally, the bulk of the evidence supports a news-media-stimulated, agenda-setting effect.

Studies of political speeches have variously examined the verbal content of the speeches, the nonverbal content, and presentation styles of mediated

(e.g., televised) speeches. For example, Hawkins (2009) examined “populist” discourse in political speeches from leaders around the world. There is a long history of applications of CATA techniques to political message content such as party platforms and political speeches (e.g., Pennings & Arnold, 2008; Suzuki, 2009). The Manifesto Project has developed a coding protocol and has compiled an online database of 3,611 political manifestos from 623 elections across 55 nations (“Project Description,” n.d.), from which hundreds of publications have been derived. CATA has been applied, as with Bligh and Robinson’s (2010) unique application of Diction to archived speeches by Gandhi, demonstrating the prevalence of eight constructs measuring “charismatic” content. And Renshon (2009) used the versatile Profiler Plus computer-coding platform to apply an “operational code” scheme, which analyzes verb usage, to compare public speeches and private communications by U.S. President John F. Kennedy during the summer of 1962, finding strong similarities between the two types.

Dozens of content analyses have looked at political advertising, making it the most-studied aspect of campaign politics. For example, Brader (2006) executed a broad-based content analysis of over 1,500 political TV ads, measuring over 100 variables per ad, including negativity appeals, informational appeals, and a host of nonverbal/visual elements denoting emotional tone. Kaid and Johnston (2001) developed a comprehensive “videostyle” analysis, focusing on a political ad’s verbal content, nonverbal content, and film/video production techniques (p. 27), including both substance and form variables. Their analysis of all TV ads for U.S. presidential candidates from 1952 to 1996, collected from the Political Commercial Archive at the University of Oklahoma, found that the commercials of winners were more likely to use logical appeals, emphasize competency and values, attack the record of an opponent, use testimonials rather than anonymous announcers, use more special effects, and be of shorter length.

One focus of the content analyses of political advertising has been that of negative, or “attack,” ads (Kaid & Bystrom, 1999). Motivated by effects research that has shown both negative and positive shifts in opinions toward the negative ad target as a result of exposure (Lin, 1996; Schenck-Hamlin, Procter, & Rumsey, 2000), content analyses have looked at the frequency of the appearance of negative political ads (Finkel & Geer, 1998) as well as the news coverage of the controversial ads (Kaid, Tedesco, & McKinnon, 1996). This focus has been extended to studies of online political promotions, such as Druckman, Kifer, and Parkin’s (2010) study of negative appeals on congressional candidate web sites.

## Violence in the Media

When we try to explain content analysis to a layperson, we often fall back on the example of media violence: “You know, those studies that have

looked at how much violence there is on TV” (or in music videos, or in video games, depending on what decade it is). This is the application of content analysis that has received the most public attention over the years, from the early Payne Fund Studies of motion pictures and youth in the 1930s (Dale, 1935) through the storm of attention given to aggressive TV content and its potential effects on children in the 1960s and 1970s (Lange et al., 1969; Comstock & Rubinstein, 1972) to subsequent multiyear efforts to track the level of violence on TV (Gerbner et al., 1980; Greenberg, 1980; *National Television Violence Study*, 1997). Thousands of studies have examined some aspect of violent or aggressive behavior in media content or its effects on adult and child audiences (see Potter, 2008). “Milestones in Content Analysis History,” available on *The Content Analysis Guidebook Online* (CAGO), highlights several of the major initiatives that not only addressed the problem of violence in American society but also contributed in meaningful ways to the repertoire of content analysis *methods* (Gerbner, Signorielli, & Morgan, 1995; Greenberg, 1980).

The research on violent media content may be seen as progressing from raw number counts (Dale, 1935) to a consideration of a variety of types of aggressive behavior, including verbal aggression (Lange et al., 1969; Greenberg, 1980, who found about half of all aggressive behaviors on TV to be verbal) to a careful consideration of the *context* of the aggression (Kunkel et al., 1995, Potter & Ware, 1987). As Kunkel et al. (1995) point out, industry executives argue that “the meaning and impact of violence in a film such as *Schindler’s List* is quite different from that of *Terminator II*” (p. 285). Their multiyear content analysis of television included such contextual variables as the nature of the perpetrator and of the target (e.g., human vs. animal, gender, age, ethnicity, “good” vs. “bad”), the reasons for the violence (e.g., personal gain, anger, retaliation), the means or method of violence (e.g., firearms, so-called natural means), the extent of the violence (how many behavioral acts in a violent interaction), the graphichness of the violence (e.g., presence of blood and gore, long shot vs. close-up), the realism of the violence (fantasy, fiction, reality), rewards and punishments associated with the violence, consequences of the violence (amount of harm), and presence or absence of humor in the violent scene.

Such analyses have been applied to other media and forms of entertainment, adding to the mix studies of aggressive behaviors in movies (e.g., Neuendorf et al., 2010; Spicer, 2012), music videos (see Smith, 2005, for a review), pornography (e.g., Bridges et al., 2010), sports (e.g., Tamborini et al., 2005), web sites (e.g., Gossett & Byrne, 2002), and video gaming.

Pieper, Chan, and Smith (2009) summarized the content analysis research on violence in video games, raising questions of sampling, time frames, and units of analysis. And since content analyses of aggressive behaviors are motivated by concern over potential effects on audience members, the basic motivation of the aggression is of interest; Sherry (2001) separates video game violence into human, fantasy, and sports violence, each with different audience outcome potentials.

## Gender Roles in the Media

Perhaps no substantive area has been more frequently studied across all the mass media than that of the roles of males and females. The journal *Sex Roles* is a prime venue for content analyses with a gender focus. A limited number of content analyses have looked at gender role behaviors among real people—comparing males and females with regard to dreams, memories, perceptions of the world, interrupting behaviors, use of a variety of language dimensions, type of explanatory style, the use of Facebook profile photographs, the use of online cancer communities, and physician verbal behaviors, for example (Barrett & Lally, 1999; Buchanan & Seligman, 1995; Domhoff, 1999; Ginossar, 2008; Greener & Crick, 1999; Hum et al., 2011; Lance, 1998; Marche & Peterson, 1993; Pennebaker, 2011; Roter et al., 2002). This leaves the many mass media-focused content analyses with relatively few bases for real-world comparisons.

An amazing variety of media form and content types have been examined for their gender role portrayals. Studies have compared male and female roles and behaviors for domestic or international content in television, film, news coverage, magazines, radio talk, textbooks, children's books, advertising of all types, comics, video games, software, rock and roll music, music videos, slasher films, birth congratulatory cards, and even postage stamps. Generally, the findings confirm a message environment of androcentrism (i.e., with males heavily overrepresented in sheer numbers and routinely given more important roles) and sex stereotyping (i.e., with significant and often predictable differences between male and female characterizations, a robust finding that has endured over time; Barner, 1999; Chappell, 1996; Drewniansy, 1996; Emons, Wester, & Scheepers, 2010; Gerding & Signorielli, 2014; Gottschall et al., 2008; Greenberg, 1980; Hether & Murphy, 2010; Kalis & Neuendorf, 1989; Lemish & Tidhar, 1999; Low & Sherrard, 1999; Mastro & Ortiz, 2008; Michelson, 1996; Neuendorf, 2011; Ogletree, Merritt, & Roberts, 1994; Watkins, 1996; Weaver, 1991).

When compared to females, males have been characterized as more active on birth cards (Bridges, 1993), less argumentative on radio talk shows (Brinson & Winn, 1997), less likely to be portrayed as sex objects in video games (Dietz, 1998), and older in films of the 1930s, 1940s, and 1990s (Smith, 1999). However, Signorielli and Bacue (1999) found some changes in occupational depictions for women from the 1960s through the 1990s. Extending their examination beyond content alone (using an integrative approach), Lauzen, Dozier, and Hicks (2001) found involvement of female TV executives and creative personnel to be predictive of the use of powerful language patterns by female characters. On the other hand, Lindner (2004) studied women's images in magazine ads between 1955 and 2002, using a coding scheme adapted from Erving Goffman's work on gender roles in the 1970s, and found greater gender stereotyping in the female-focused *Vogue* magazine than in the general-interest magazine *Time*. As noted earlier, in a



study of television and newspaper sport coverage, Kian, Mondello, and Vincent (2009) failed to confirm numerous expected differences in the descriptions of male and female basketball players, finding significant differences for only between physical appearance/attire (more mentions for men) and psychological/emotional strength (more mentions for women).

Busby's (1975) review of female images in the media is an excellent, comprehensive historical source. Walsh and Ward (2008) reviewed adolescent gender roles across a range of six media since 1950, finding "limited, if any, change . . . over the past thirty to fifty years" (p. 151). More updates may be found in two special issues on content analysis in the journal *Sex Roles* (June 2010 and February 2011; see Rudy, Popova, & Linz, 2010, 2011). In these two issues, 18 research articles are collected, examining the underrepresentation, sexualization, subordination, and role portrayals of the genders in venues including video games, music videos, slasher films, top-grossing films, international TV advertising, toy commercials, local TV news, and Spanish newspapers (Rudy, Popova, & Linz, 2011). In sum, the collection confirms the underrepresentation of women across media and across the globe. Further, stereotypical portrayals continue to dominate (Collins, 2011). For example, Wallis (2011) concluded that in televised music videos, women continue to be shown as fragile, subordinate, and sexually suggestive, while men are generally shown as aggressive rather than sexual.

One very particular application of gender-based content analysis has been the examination of *face-ism*, the tendency of a photo to reveal more of the subject's face or head than body. Associated with audience perceptions of dominance and positive affect (Levesque & Lowe, 1999; Zuckerman, 1986), higher face-ism has been generally granted to males over females in the mass media (Archer et al., 1983; Copeland, 1989; Sparks & Fehlner, 1986). This finding has been extended to an online imagery (e.g., Szillis and Stahlber's [2007] study that found male professors and members of the Parliament of Germany to have greater face-ism) and confirmed in studies of social media (e.g., Smith & Cooley's [2012] finding of "self-inflicted face-ism" on Facebook).

Another particular content analysis focus that has received much attention in recent decades is that of body image (weight and shape). Although increasingly applied to depictions of males in the media (e.g., Jankowski et al., 2014), body image is usually studied as a potential influence on young female audience members' self-concept (Botta, 2000; Zhang, Dixon, & Conrad, 2009). Harrison (2008) reviewed the literature on body image and eating in the media since the 1950s, concluding that "demonstrations of the thin female adolescent body ideal and the male muscular body ideal have become increasingly ubiquitous in American media" (p. 191; see also Martins et al., 2009, 2011). Research has found low-weight female characters to be over-represented and highly praised by other characters on television situation comedies (Fouts & Burggraf, 1999). And reality TV portrayals of weight loss have been found to be unrealistic (e.g., Klos et al., 2015).

## Sex and Sexuality in the Media

---

In addition to the health-related research on sexual behaviors in the media, other content analyses have examined issues of sexual behavior and of sexuality. Hetsroni (2007) conducted a meta-analysis of 25 content analyses of sexual content on American prime-time television airing between 1975 and 2004, uncovering the following overall trends: decreasing frequency of dialogue about sex, portrayals of normative heterosexual behaviors, portrayals of illegal sexual behaviors, and messages about risks and responsibilities with regard to sexual behaviors; and increasing frequency of portrayals of homosexuality (although this is apparently limited to instances of identified sexual orientation; none of the 25 studies identified even a single instance of same-sex intercourse, implied or explicit). In an analysis of the 2005–2006 TV/cable season, however, Netzley (2010) identified instances of “sexually active” gay characters; 7.5% of all characters on prime time were found to be gay/bisexual, and they were more likely to be shown in “sexual situations” than were straight characters.

Greene et al. (2011) analyzed both talk about sex and portrayed sexual behaviors in television programming, finding more sexual content on fictional rather than reality television programming. Stern and Brown (2008) provide a review of research relevant to what types of sexual media content adolescents have been exposed to over the past 50 years, providing a good overall history of sex in various media. While portrayals of sexual behaviors have increased over time, responsible discourse about sexual activity has not kept pace. Hust, Brown, and L’Engle (2008) studied content popular with adolescents (i.e., ages 12–14) in four media (TV, magazines, music, and movies), finding that less than half of 1% of the content included information about or depictions of sexually healthy behavior.

Content analyses of explicit sexual media content—that is, pornography—form a small but important portion of the literature on sexuality in the media. The studies range from the works during the 1970s and 1980s that first confirmed the strong presence of violence against women in pornographic films (e.g., Cowan et al., 1988; Prince, 1987) to more contemporary updates, such as Bridges et al. (2010) and Sun et al. (2008), both of which confirm a continuing high level of verbal and physical aggression in pornography, with objectification and victimization of women still quite common.

## Minority Portrayals

---

A sizable amount of research has been devoted to the systematic content analysis of the images of racial and ethnic minorities in American media. *The Howard Journal of Communications* is a prominent outlet for research on racial differences in both real-life communication behaviors and mediated portrayals.

Historically, the majority of content analyses on minority images in U.S. media have examined African American portrayals, and they include several

important reports and volumes that provide an excellent first look at some of the landmark content analyses in this area (Dates & Barlow, 1990; MacDonald, 1992; Poindexter & Stroman, 1981; Stroman, Merritt, & Matabane, 1989–1990; U.S. Commission on Civil Rights, 1977, 1979). Prior to 1970, images of Black Americans emphasized domestic and submissive roles (MacDonald, 1992). Even as gains were made in entertainment television portrayals (e.g., *The Cosby Show*), news images still emphasized a menacing African American criminal element (Atkin & Fife, 1993–1994; Barber & Gandy, 1990; Dates & Barlow, 1990; Dixon & Linz, 2000; Entman, 1992). And although subsequent content analyses tended to find that Blacks were no longer underrepresented in advertising, news, and entertainment content, findings continued to confirm relatively homogenous, stereotypical role portrayals (e.g., Keenan, 1996b; Matabane & Merritt, 1996; Plous & Neptune, 1997; Wilkes & Valencia, 1989). More contemporary studies have found “subordinate” roles in advertising for Blacks (Hollerbach, 2009), limited portrayals in business and work-related settings for Black males in advertising (Bailey, 2006), and negative descriptions of personal characteristics of African American players in intercollegiate sports commentary (Rada & Wulfemeyer, 2005).

Fewer studies have examined the status of Latinos/Latinas, Asian Americans, or Native Americans in entertainment, commercial, or news media content (e.g., Greenberg et al., 1983; Singer, 1982, 1997; Taylor, Lee, & Stern, 1996; Taylor & Stern, 1997; Wilkes & Valencia, 1989), although this is changing (e.g., Dixon, 2016, in press; Dixon & Williams, 2015). Stroman and Dates (2008) reviewed the extant literature, concluding that the evidence points toward continued severe underrepresentation for these groups. And stereotypical portrayals maintain (e.g., Buzinde, Santos, & Smith, 2006; Mastro & Behm-Morawitz, 2005), even when those portrayals are generally positive (e.g., the Asian American as the hard-working model minority; Taylor & Stern, 1997). Further, attention has begun to focus on portrayals of Arabs and of Muslims, two groups for which “otherness” is a common theme of their often negative media portrayals in the 21st century (e.g., Dixon & Williams, 2015; Powell, 2011; Saeed, 2007; Trevino, Kanso, & Nelson, 2010).

Like the research on women’s roles in the media, studies on racial and ethnic minorities have encompassed a variety of media sources, including textbooks, movies, news coverage, magazines, pornography, reality-based crime shows, direct-mail advertising, editorial cartoons, jokes, and viewer ratings of YouTube videos (Correa, 2010; Cowan & Campbell, 1994; Kopacz & Lawton, 2011; Oliver, 1994; Spencer, 1989; Spicer, 2012; Stevenson & Swayne, 1999; Zurbriggen & Sherman, 2010).

---

## Movies

Only in recent years has a critical mass of quantitative content analyses of narrative films accumulated. This is at first glance puzzling, given that some

of the earliest content analyses with reasonably sound methods were done on American movies. Motivated by a concern over potential effects of movie content on children, the Payne Fund Studies of the 1920s and 1930s combined laboratory studies, field experiments, surveys, and content analyses that examined themes and social values prevalent in movies at the time (see “Milestones in Content Analysis History” at the *Content Analysis Guidebook Online*). However, it would be many decades before content analyses of film would become common. Two possible reasons for this “late arrival” of movie content analyses are (a) the focus within the scholarly film literature on critical/cultural approaches, to the near exclusion of other methods, and (b) the unavailability until recent decades of readily accessed archived film content.

Content analysts seem to be making up for lost time. There seems to be a recognition that narrative films are an enduring set of cultural artifacts, potentially important sources of cultural learning (Beckwith, 2009). In addition to the semiautomated analyses of the form features of color distribution and editing rhythm described in Chapter 7 (e.g., Anderson & O’Conner, 2009; Cutting, DeLong, & Nothelfer, 2010), human coding of films has been expanding in popularity. Character portrayals have been a common focus for content analyses, with particular emphasis on women’s images throughout film history (e.g., Capwell, 1997; Liebler, Jiang, & Chen, 2015; Neuendorf et al., 2010; Smith, 1999). Beyond this, the range of study types is rather amazing. Content analysis of film has been used to examine the following: Auteur theory (Janstova, 2006); the changing values of protagonists in top commercial movies (Beckwith, 2009); the portrayal of schizophrenia in contemporary films (Owen, 2012); the prevalence of sexual and violent imagery in movie previews (Oliver & Kalyanaraman, 2002); language choice by male and female screenwriters (Pennebaker, 2011); the representation of age and gender in popular films (Lauzen & Dozier, 2005); the portrayal of stepfamilies in the movies (Leon & Angst, 2005); the characteristics of romantic relationships in romantic comedies (Johnson & Holmes, 2009); portrayals of inter- and intraracial relationships involving African Americans and European Americans (Beeman, 2007); “quotable” movie quotes (using natural language processing; Danescu-Niculescu-Mizil et al., 2012); and product placement in films (a number of studies have examined this; e.g., Cassady et al., 2006; El Damanhoury, 2015).

Discrete behaviors in movies have been scrutinized. In perhaps the most comprehensive longitudinal examination, Jamieson et al. (2008) tracked “health risk behaviors”—violence, sex, smoking, alcohol consumption, and suicide—in top-grossing movies from 1950 to 2004. Additionally, studies have looked at smoking and drinking alcohol (Dalton et al., 2002; Tickle, Beach, & Dalton, 2009); food-related activities and exercise (Bell et al., 2005); swearing in movies over a 26-year period (Cressman et al., 2009); sexual and violent behaviors by gender (Bleakley, Jamieson, & Romer, 2012); indirect aggression in Disney animated features (Coyne & Whitehead, 2008);

prosocial behaviors in Disney animated features (Padilla-Walker et al., 2013); risky driving in popular action movies (Beullens, Roe, & Van den Bulck, 2011); and the treatment of Native Americans in the Westerns of John Ford (Spicer, 2012). Content analyses have also contributed to the emerging mini-literature on the “Disney princesses” phenomenon (e.g., England, Descartes, & Collier-Meek, 2011).

Further, an interesting trend has developed, focusing on elemental aspects of film editing. Initiatives around the world are dedicated to the compilation of shot data for film and television, although these initiatives are not labeled content analyses, and they do not seem to use coder training or reliability assessment. Cinemetrics, developed by film scholars Yuri Tsivian and Gunars Civjans (n.d.), is a sort of “crowd-sourced content analysis” project that has to date collected shot-length data from an incredible 15,790 films. The Institute for Research and Innovation of the Centre Pompidou has developed Lignes de Temps (Time Lines), software that facilitates shot-by-shot analysis of films (see Butler, 2014). And Salt (2005, 2009, 2011) has conducted longitudinal analyses of movie shot lengths, shot types, and selected camera movements and editing techniques.

---

## Music

The lyrics of popular songs have been the subject of content analyses over the years (e.g., Dodds & Danforth, 2010; Dukes et al., 2003; Langdon, 2012; Pettijohn & Sacco, 2009; Rothbaum & Tsang, 1998; Rothbaum & Xu, 1995) as indicators of social trends or for cross-cultural comparisons, often linking the analyses of lyrics to documented changes in societal indicators such as the economy (see the Zullo, 1991, example in Chapter 2).<sup>5</sup>

Less often, the nature of the music itself has been content analyzed (e.g., melodies, instrumentations, and motifs; Narmour, 1996; Simonton, 1994, 2003b, 2010). But the greatest number of investigations relevant to popular music have been content analyses of music videos, generally addressing a concern over violent, sexist, and race-typed images (e.g., Aubrey & Frisby, 2011; Baxter et al., 1985; Conrad, Dixon, & Zhang, 2009; Jones, 1997; Kalis & Neuendorf, 1989; Sherman & Dominick, 1986; Wallis, 2011).

---

## Sentiment Analysis

Introduced in Chapter 5, sentiment analysis is a comparatively recent addition to the complement of techniques within content analysis (Liu, 2010, 2012; Pang & Lee, 2008). Sentiment analysis is most commonly applied to online social media posts by members of the general public in order to gauge public evaluations of products, current issues, or other subjects of interest.

Used for both academic and applied purposes, sentiment analysis has grown rapidly in use, fed by the availability of textual “big data” online, an unprecedented collection of user-generated content. Systems that execute sentiment analyses range from those that execute basic word searches using positive and negative preset dictionaries (e.g., applications of LIWC or SentiStrength; see Resource 1) to those that incorporate neural network models that have been trained to decipher grammatical construction (e.g., Stanford University’s sentiment analysis initiative, see <http://nlp.stanford.edu/sentiment>).

Sentiment analysis has been applied to, among other types of user-generated content, personal blogs (Neviarouskaya, Prendinger, & Ishizuka, 2009), instant messaging (Neviarouskaya, Prendinger, & Ishizuka, 2007), web forums (Abbasi et al., 2008), MySpace comments (Thelwall, Wilkinson, & Uppal, 2010), Twitter content (Thelwall & Buckley, 2013), and Facebook messages (Ortigosa, Martin, & Carro, 2014). Importantly, sentiment analysis has been proposed as a proxy for traditional public opinion polling (e.g., Ceron et al., 2014).

The commercial applications of sentiment analysis are usually proprietary, although some result in news releases that draw our attention to their existence. For example, in 2014 Facebook reportedly manipulated users’ news feeds and measured the response using sentiment analysis; their manipulation was the subject of much scrutiny (Wohlsen, 2014).

## Academic Literatures

---

Content analysis has joined the repertoire of techniques for reviewing and summarizing bodies of academic research. More concertedly objective and quantitative than the venerable literature review and more adaptable than the statistical summaries of findings of meta-analysis, content analysis affords the opportunity to examine past studies for their theories, concepts and constructs, methods, reportage, and, really, anything else the scholar may wish to tap.

The variety of content analyses of academic literatures has included investigations of the following: 30 years of research articles on college drinking (Broughton & Molasso, 2006); e-government research in two journals and one conference series (Heeks & Bailur, 2007); the content of 15 journals in selling and sales management over a 20-year period (Williams & Plouffe, 2007); articles on business intelligence in 10 leading IT journals over a 10-year period (Jourdan, Rainer, & Marshall, 2008); the concepts of race, ethnicity, and racism as appearing in medical-anthropological research journals over a 25-year period (Gravlee & Sweet, 2008); changes in the major topics discussed at the first 20 annual conferences of the North American Serials Interest Group (NASIG; an organization for librarians, publishers, and related professionals; Garner, Davidson, & Williams, 2008); the distribution of articles on seven emergent subfields of urban studies as appearing in the *Journal of Urban Affairs* over a 9-year period (Bowen, Dunn, & Kasdan,



2010); the treatment of bisexuality in the medical literature over a 20-year period (Kaestlea & Ivory, 2012); the content of academic articles on group counseling involving five U.S. minority groups over a 25-year period (Stark-Rose et al., 2012); trends in topic coverage over a 40-year period in the *Journal of Moral Education* (Lee & Taylor, 2013); contents and methods for articles on Internet marketing in top IT and marketing journals over an 18-year period (Corley, Jourdan, & Ingram, 2013); the focus, pedagogical foundations, and methodologies apparent in articles on pedagogy in counselor education over a 10-year span (Minton, Morris, & Yaites, 2014); and theories, topics, media, and methods for 30 years of advertising research articles in 17 top journals (Kim et al., 2014).

## Commercial, Governmental, and Other Client-Based Applications of Content Analysis

As the popularity of content analysis as an academic research tool has grown, there has been a corresponding development in interest in applying content analysis to the needs of business, government, and other nonacademic entities. These include instances of the public sector and private-sector funding of content analysis research conducted by academics, the strictly-for-hire content analysis work of academic researchers, and the use of content analysis in the private sector.

The private funding of academically based content analysis is a longstanding tradition, from the Payne Fund Studies of movies funded by the Payne Foundation and executed by researchers at six universities in the 1920s to the National Television Violence Study funded by the National Cable Television Association and the Sex on TV project funded by the Kaiser Foundation, both in the 1990s. Government funding has also bolstered a fair number of large-scale content analyses based at universities, such as Project CASTLE (Children and Social Television Learning), funded in the 1970s by the U.S. Office of Child Development of the Department of Health, Education, and Welfare (Greenberg, 1980) and the Content and Effects of Alcohol Advertising project, funded by the U.S. Bureau of Alcohol, Tobacco, and Firearms in the late 1970s.<sup>6</sup>

Public relations, advertising, marketing, and public opinion polling firms have sometimes included content analysis as one of the services they have offered to clients over the years (Lindenmann, 1983; Stone, 1997). Lindenmann describes the in-house research efforts of AT&T, which in the 1970s began a program of content analysis of newspaper coverage of the corporation to gauge PR effectiveness. Macnamara (2005) proposes at least two applications of content analysis in the PR arena: (1) strategic planning, acquiring insights and intelligence via “issues tracking (environmental scanning), competitor analysis, and trend identification”; and (2) evaluation research to measure the effectiveness of an organization’s PR communications via media, including

“audience reach, messages communicated, ‘share of voice,’ and benchmarking its profile against competitors or in its sector” (p. 21).

CATA procedures are attractive to commercial clients. CATA has been applied to transcripts of focus groups by marketing firms (Camden & Verba, 1986; Stone, 1997) and to news coverage by private consultants (Fan & Bengston, 1997; McCarty, 2001). Psychological diagnostic CATA techniques are often used by practitioners in a commercial context (e.g., via PCAD; see Resource 1). The CATPAC semantic network computer program has been used extensively for consulting purposes (Salisbury, 2001; [http://www.galileo.com/N\\_client.asp](http://www.galileo.com/N_client.asp)). In some instances, content analysis is used by media institutions for self-monitoring their own messages (e.g., Wurtzel & Lometti, 1984). Basic readability measures are used to set and check writing standards by newspapers, magazines, and wire services (Danielson et al., 1992; Severin & Tankard, 1997).

Unfortunately, many of the purely commercial applications of content analysis are proprietary and unavailable for us to learn from. There is anecdotal evidence that commercial content analysis studies do not always follow the good-science guidelines presented in this book. Industry standards for content analysis, such as those established for polling by the American Association for Public Opinion Research, are not yet in place. There is a need for the establishment of self-monitoring standards for commercial content analysis, and these should include a call for open and full reportage of methods.

## Legal Applications of Content Analysis

Hall and Wright (2008) have discovered and summarized numerous content analyses of judicial opinions that have been executed over the years. Dating back to a 1957 analysis of Supreme Court decisions, these content analyses involve systematic coding of legal rulings, providing a quantitative touchstone for case law interpretations, looking at factors that drive judges to certain decisions or trying to use past decisions to predict future decisions of the court. Although the 134 content analyses reviewed by Hall and Wright have been published in academic outlets, there are clear real-world implications for their findings.

Similarly, Liu et al. (2015) demonstrated the utility of content analysis for the comprehensive and objective documentation of congressional testimony. They analyzed all 1,350 testimonies from scientist witnesses during congressional global warming and climate change (GWCC) hearings over a period of 39 years from 1969 to 2007, finding the vast majority of the scientists expressing the views that GWCC is real (86%), that GWCC is at least to some degree caused by human activity (78%), and that proactive policy should be enacted (95%).

Client-focused content analyses have assisted law firms in at least a couple of ways. First, CATA analysis and human coding of news coverage has been

used for change of venue petitions, as in the set of civil court cases filed against former Cleveland Browns owner Art Modell when he moved the pro football team to Baltimore in 1996 (McCarty, 2001).<sup>8</sup> Second, content analyses have been employed to analyze news coverage over time of key topics related to lawsuits, as in the 2012 Toyota unintended acceleration (UA) case, for which human and CATA analyses tracked the volume of coverage and the variety of ways in which news stories about UA were framed.<sup>9</sup>

---

## Future Directions

The first edition of this book posed predictions for content analysis 100 years hence. Well, many of those predictions made in 2002 have actually come to pass in little more than a decade!

As predicted, private messages are virtually all “archived” in some fashion online, and it has become difficult to keep any message fully private in America (the NSA has seen to that). It was predicted that for receiver-based content analyses—that is, those concerned with the effects of messages—there would be individualized, not aggregated, content analyses; it is already becoming feasible and popular to tap receivers’ exposure patterns and then conduct content analyses on the actual content known to be consumed. In 2002, the first edition of this text invoked the idea of a future “Big Sibling” world and noted it was similar to M.I.T. scientist Ed Fredkin’s notion of reality as composed of information and the universe as a giant data processor (Wright, 1988). “In this universe of information, there will be no such thing as sampling and inferential statistics. All information will be kept and analyzed. It will be an all-census, longitudinal data world,” I opined (Neuendorf, 2002, p. 212). In the era of big data, we are at the cusp of this reality.

The exponential rate of change makes future predictions precarious. I guess I would repeat some predictions that have not yet come true—for example, content analyses attempting to infer to the source in the future will routinely be in real-time, including individualized sentiment analysis; individuals may have anatomically embedded chips that will transmit data about their activities to online databases that will be automatically content analyzed (we are close to that now); metadata will sometime in the future be recognized for its great content analytic potential, and more comprehensive and standardized systems for its use will be established; advancements in artificial intelligence will inexorably take control of content analyses out of the hands of humans; the content analysis of thoughts and dreams will be the new frontier for content analysts of the future.

Standards for content analysis execution and reportage are being upheld to a much greater degree than they were in 2002. The first edition of this book called for higher, and more widely accepted, standards for the execution of content analysis, both academic and commercial. This included full reportage of methods for purposes of replicability and comparative analyses. While