Researching Interactive Communication Behavior

A Sourcebook of Methods and Measures

C. Arthur VanLear ▪ Daniel J. Canary EDITORS

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

12

CODING OBSERVED INTERACTION

ALAN L. SILLARS AND NICKOLA C. OVERALL

In this chapter, we discuss practical and conceptual issues when coding observed communication. At first glance, the process can seem straightforward: One selects a coding system, trains coders to use the manual, and checks reliability. However, coding requires more than mechanically applying categories or ratings to message units. Coding is a form of message interpretation, analogous to what happens in all communication (Folger, Hewes, & Poole, 1984). Coders, like participants in communication, apply interpretive rules to discourse and nonverbal behavior in order to discern meaning, either conventional meaning or meaning specific to observer or participant goals. In observational coding, as in everyday communication, standardized coding rules promote shared meaning (i.e., reliability) but do not remove all ambiguity (Sillars & Vangelisti, 2006). Coders must improvise when interpreting novel or ambiguous examples, drawing on their own experience and anticipating how others would view the same message. Coding is also an exercise in selective perception. Because messages are multifunctional (Sillars & Vangelisti, 2006) and have different levels of meaning (e.g., content vs. relational), the same interaction can be coded many ways that do not inherently compete. Coding methods selectively highlight functions of communication (e.g., persuasion or support), levels of analysis (e.g., molar vs. molecular), intended meanings (e.g., observer vs. participant), structural properties (e.g., base rates vs. sequential structure), and so forth. Thus, many alternative ways of coding exist that may be appropriate (or not), depending on one's purpose and perspective.

Our experience with observational coding stems mostly from research on couple and family conflict. We draw on this experience to ground discussion of general issues in coding. Conflict is one of the most researched aspects of family communication (Sillars & Canary, 2013) and an area with a long tradition of observational work. Whereas another chapter provides a review (see D. Canary, this volume), we cite conflict coding methods selectively to illustrate issues, options, and trade-offs when conducting any form of interaction analysis.

Conceptual Foundations of Observational Coding

Observational coding typically involves coders' independently categorizing or rating the verbal and nonverbal content of a recorded interaction according to specified protocols and coding schemes. Coding yields a systematic record of ongoing communication, albeit a selective one structured by researcher assumptions and theories. As Krippendorff (2004) stressed, inference is inherent to content analysis of communication, because the outward (physical) features of messages have no meaning of their own; messages acquire "content" only when people engage them conceptually. Even automated coding performed by computers rests on theories of programmers about how humans read and respond to messages (Krippendorff, 2004). Coding supplies content by filtering, segmenting, and highlighting aspects of communication that have meaning relative to one's purpose and conceptual framework. Of necessity, the process highlights certain features while disregarding many others. Moreover, *interaction analysis* (i.e., content analysis of free-flowing

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

conversation) is especially selective. The verbal, vocal, and kinetic activities people carry out while speaking and listening are so complex and information dense per unit of time that formal analysis cannot presume to yield more than partial understanding (Street & Cappella, 1985, p. 4).

Given the interpretive and selective nature of coding, trade-offs occur when deciding to adopt a coding system, adapt one, or invent one's own. Well-studied aspects of communication, including most topics in this book, have already spawned multiple systems. It is clearly more efficient to use an existing system than to begin from the ground up. The proliferation of coding schemes also complicates synthesis of results, leading some authors to even call for a moratorium on the development of new methods (Kerig, 2001). On the other hand, adopting a coding scheme means buying into particular assumptions about what message features are important and what they signify. Thus, well-established coding options are not all purpose. Bakeman and Gottman (1997) commented that borrowing a coding scheme can feel like "wearing someone else's underwear" (p. 15), because coding represents a theoretical act originating within the confines of a particular research program.

Research on couple conflict illustrates connections between coding methods and researcher perspectives. Box 12.1 reports categories from familiar coding schemes for couple conflict, including the Marital Interaction Coding System (MICS-IV), Kategoriensystem für Partnerschaftliche Interaktion (KPI), Couples Interaction Scoring System (CISS), and Verbal Tactics Coding Scheme (VTCS). Box 12.2 reports similar codes from two rating systems: the Conflict Rating System (CRS) and Communication Strategies Coding Scheme (CSCS). (Categorical codes and ratings are discussed further under "Forms of Coding.") Collectively, the systems share much in common. Systems used to code couple conflict tend to reflect two broad dimensions: valence and directness (see Overall, Fletcher, Simpson, & Sibley, 2009; Sillars & Canary, 2013). The valence dimension is explicit in systems that collapse into positive-negative supracategories (KPI, CRS, CSCS); however, all of the coding systems have been used to operationalize positive-negative communication. Directness is reflected in engagement versus avoidance of conflict (e.g., the demand and withdraw subscales of the CRS), along with direct and indirect influence attempts (as in the CSCS). The coding systems in Boxes 12.1 and 12.2 are also similar in what they omit. That is, they foreground relational aspects of conflict at the expense of other potentially important processes, for example, bargaining tactics (Putnam & Jones, 1982) and argument structure (see Seibold & Weger, this volume). Thus, the coding schemes are well suited to research on valence and directness of conflict communication but disregard many other potentially important features.

Box 12.1
Categorical Coding Systems for Couple Conflict

Marital Interaction Coding System (Heyman, Weiss, & Eddy, 1995)

Blame (criticize, mind-read negative, putdown, turnoff)

Description (problem description, internal and external)

Dysphoric Affect

Facilitation (assent, disengage, humor, mind-read positive, positive touch, paraphrase/reflect, question, smile/laugh)

Invalidation (disagree, disapprove, deny responsibility, excuse, noncomply)

Irrelevant (unintelligible talk)

Propose Change (compromise, negative and positive solution)

Validation (agree, approve, accept responsibility, comply)

Withdrawal

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Kategoriensystem für Partnerschaftliche Interaktion (Hahlweg, 2004)

Positive Verbal

Self-Disclosure (expression of feelings, wishes, attitudes, or behavior)

Positive Solution (constructive proposal, compromise suggestions)

Acceptance of the Other (paraphrase, open question, positive feedback, understanding, agreement)

Neutral Verbal

Problem Description (neutral description, neutral questions)

Meta Communication (clarifying requests, related to topic)

Rest (inaudible or does not fit other categories)

Listening

Negative Verbal

Criticize (devaluation of partner, specific criticism)

Negative Solution (destructive solution, demand for omission)

Justification (excuse own behavior, deny responsibility)

Disagreement (direct disagreement, yes-but, short disagreement, blocking off)

Couples Interaction Scoring System (Gottman, 1979)

Content Codes

Problem Information or Feelings About a Problem

Mindreading

Proposing a Solution

Communication Talk

Agreement

Disagreement

Summarizing Other

Summarizing Self

Nonverbal Behavior

Positive (face, voice, and body cues such as smiling, warm voice, touching)

Negative (face, voice, and body cues such as frown, cold voice, inattention)

Neutral (absence of positive or negative nonverbal cues)

(Continued)

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

(Continued)

Verbal Tactics Coding Scheme (Sillars, 1986)

Denial and Equivocation (direct or implicit denial, evasive replies)

Topic Management (topic shifts, topic avoidance)

Noncommittal Remarks (noncommittal statements and questions, abstract or procedural remarks)

Irreverent Remarks (friendly joking)

Analytic Remarks (descriptive, disclosive, or qualifying statements; soliciting disclosure or criticism)

Confrontative Remarks (personal criticism, rejection, hostile imperatives, hostile jokes, or questions, presumptive attribution, denial of responsibility)

Conciliatory Remarks (supportive remarks, concessions, acceptance of responsibility)

Source: Author.

Box 12.2
Rating Systems for Couple Conflict

Conflict Rating System (Heavey, Lane, & Christensen, 1993)

Demand Subscale

Discussion (tries to discuss a problem, is engaged and emotionally involved)

Blames (blames, accuses, or criticizes; uses sarcasm or character assassination)

Pressures for Change (requests, demands, nags, or otherwise pressures)

Withdraw Subscale

Avoidance (hesitating, changing topics, diverting attention, or delaying discussion)

Withdraws (withdraws, becomes silent, refuses to discuss topic, looks away, disengages)

Positive Subscale

Negotiates (suggests solutions and compromises)

Backchannels (shows listening through positive minimal responses)

Validates Partner (indicates verbal understanding or acceptance of partner's feelings)

Positive Affect (expresses caring, concern, humor, or appreciation)

Communicates Clearly (expresses self in a way that is easy to understand)

Negative Subscale

Expresses Critical Feelings (verbally expresses hurt, anger, or sadness directed at partner)

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Interrupts

Dominates Discussion (dominates, tries to take control of the discussion)

Negative Affect (verbal or nonverbal anger, frustration, hostility, hurt, or sadness)

Communication Strategies Coding Scheme (Overall, Fletcher, Simpson, & Sibley, 2009)

Negative-Direct

Coercion (derogate partner, indicate negative consequences for partner, display negative affect, accuse and blame partner)

Autocracy (insist or demand, talk from a position of authority, invalidate partner's point of view, take a domineering and/or nonnegotiative stance)

Negative-Indirect

Manipulation (attempt to make partner feel guilty, appeal to partner's love and concern)

Supplication (use emotional expression of hurt, debase self and/or present self as needing help, emphasize negative consequences for self)

Positive-Direct

Rational Reasoning (use and seek accurate information, use logic and rational reasoning, explain behavior or point of view in a way the partner would find reasonable)

Positive-Indirect

Soft Positive (soften persuasion attempts, encourage partner to explain point of view and express feelings, acknowledge and validate partner's views, be charming and express positive affect)

Source: Author.

Do not copy, post, or distribute

Despite broad similarities, the coding schemes in Boxes 12.1 and 12.2 also reflect important differences that stem from research goals and observational contexts. Some coding schemes originating in clinical psychology, such as the MICS, KPI, and CISS, were designed to isolate communication skill deficits of unhappy couples as a basis for couple therapy. Early studies in this tradition conceptualized communication according to social learning principles, as contingent patterns of positive and negative behavioral reinforcement (Birchler, Weiss, & Vincent, 1975; Gottman, 1982). Thus, codes are organized and aggregated into positive and negative forms of communication, partly on the basis of how messages are presumed to affect marital outcomes. Although this division serves a purpose for behaviorally oriented therapists, others might find the approach limiting. In their dialectical critique of the satisfaction literature, Erbert and Duck (1997) chafed at the notion that interaction characteristics discriminating adjusted-maladjusted relationships can be dichotomized as positive or negative communication. In their view, the positive-negative duality reinforces an idealized view of relationships as either happy or conflicted and obscures ways that interactions may be simultaneously positive and negative.

In contrast to clinically based research, Sillars developed the VTCS with the assumption that dyadic interaction styles may have variable associations with outcomes, depending on relationship context (see Sillars & Wilmot, 1994).[1] Similarly, Overall, Fletcher, Simpson, and Sibley (2009) developed the CSCS to move past assumptions that "positive" and "negative" messages inherently

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

benefit or harm relationships by distinguishing between direct (e.g., *coercion*) and indirect (e.g., *manipulation*) influence strategies. Research using the CSCS and VTCS provides evidence that seemingly "negative" acts can sometimes help couples directly tackle relationship problems (McNulty & Russell, 2010; Overall et al., 2009).

The treatment of avoidance also differs across conflict coding schemes. Early generation coding systems in psychology (e.g., MICS, CISS; Box 12.1) primarily featured direct forms of conflict engagement (although *withdrawal* was added as a category in the fourth revision of the MICS). This reflects the main observational method, the *problem-solving paradigm*, whereby couples interact in a lab under instruction to discuss and resolve an acknowledged problem (Gottman, 1994, pp. 18–19). Although the problem-solving paradigm remains a dominant approach, later generation systems (e.g., the CRS[2]; Box 12.2) focus more on withdrawal from interaction. In contrast to research using the problem-solving paradigm, the VTCS (Box 12.1) was developed from research that allowed greater latitude for conflict avoidance and neutrality; for example, couples were instructed to discuss *potential* conflicts "until they had nothing further to say" (e.g., Sillars, Pike, Jones, & Murphy, 1984). Consequently, the VTCS distinguishes nonengagement tactics more than do other coding schemes.

Despite these contrasts, all coding systems in Boxes 12.1 and 12.2 rely on structured observation, at home or in a lab, whereby researchers prompt couples to discuss relationship issues. No doubt, naturalistic observation of conflict would reveal other forms of avoidance, such as leaving the scene, retreating to electronic devices (Heyman, Lorber, Eddy, & West, 2014), or interspersing confrontation with attention to daily tasks (Sillars & Wilmot, 1994). Observational context also affects the dimensions of communication readily observed. For example, the coding schemes in Boxes 12.1 and 12.2 contain more "negative" codes than "positive" or constructive ones. Heyman (2001) noted, "Whereas it is relatively easy to get unhappy couples to argue on command, behaviors that promote the various forms of love . . . are much more challenging to witness in the laboratory" (p. 7).

In sum, coding schemes connect to researcher assumptions, goals, and observational methods. No coding scheme can suffice for all purposes, and most require significant adaptation when there is a shift from the original context in which methods were developed.

Forms of Coding

Coding may take a variety of forms, including categorical codes, checklists, and ratings. Each approach invokes conceptual and practical trade-offs.

Discrete Coding Systems

Categorical codes. In the classic sense, coding involves classifying message units into mutually exclusive and exhaustive categories (Krippendorff, 2004). Categorical coding schemes are sometimes referred to as *micro* codes, because they code communication at the level of individual messages, whereas *macro* codes (e.g., ratings) describe longer segments of interaction (Lindahl, 2001). The CISS, KPI, MICS, and VTCS (Box 12.1) illustrate categorical coding schemes. These systems first identify a unit of observation (such as the speaking turn or thought unit[3]) and then exhaustively code these units into a fixed set of categories. Subcategories might be nested under broader categories in order to yield a more detailed description at the level of subcategories, while providing sufficient observations for quantitative analyses after collapsing codes (e.g., *blame* in the MICS-IV is a combination of *criticize*, *mindread*, *putdown*, and *turn-off*).

The primary advantages of categorical codes are their descriptiveness and flexibility. Although not nearly as fine grained as qualitative conversation analysis (CA; Robinson, 2011), categorical coding yields a more detailed record than do other forms of quantitative interaction analysis.[4] Categorical coding is also conducive to statistical analysis of sequential structure, which examines whether specific codes elicit an immediate response (VanLear, this volume). In relationship conflict, important sequences include the probability that negative codes are reciprocated by the partner (negative reciprocity) or that demand is followed by withdrawal. The categorical coding systems in Box 12.1 were developed in a period marked by influential calls to focus on the temporal organization of interaction as a way to operationalize systems thinking about relationships (e.g., Gottman, 1979; Watzlawick, Beavin, & Jackson, 1967). Categorical codes also offer flexibility in subsequent aggregation, assuming that the initial round of coding identifies more than a few categories. When detailed codes are aggregated into broad categories, the research can document how specific codes contribute to summary scores. Unfortunately, this step is often omitted when researchers report aggregate codes.

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

The time and expense of categorical coding pose a clear trade-off. For instance, trained coders need 1½ to 2 hours to analyze a 10-minute interaction using the MICS (Heyman, 2004) and even longer periods using the CISS (Notarius, Markman, & Gottman, 1983). Detailed coding of interactions requires, at minimum, an audio (and sometimes video) record, and is usually assisted by written transcripts. In addition to the time and expense of transcription, the interaction record must be *unitized*, which requires separate coder training and reliability assessment if the unit of analysis involves significant coder judgment (as with thought units). Coding itself can require difficult decisions about how to assign borderline examples to similar categories, which fatigue coders and contribute to poor reliability. Thus, as Heyman et al. (2014) noted, microanalytic coding carries a poor cost-benefit trade-off when a large number of initial categories are later aggregated into just a few (e.g., positive vs. negative communication).

One way to make coding more efficient is to apply coding schemes selectively, using only the categories of greatest relevance. For example, McNulty and Russell (2010) limited their use of the VTCS (Box 12.1) to negative (i.e., *confrontative*) codes, as their purpose was to assess longitudinal impacts of negative messages on marital satisfaction. Others have developed "rapid" coding systems, such as the Rapid Couples Interaction Scoring System (RCISS; Krokoff, Gottman, & Hass, 1989) and Rapid Marital Interaction Coding System (RMICS; Heyman, 2004), which mimic the CISS and MICS (Box 12.1) but dispense with detailed subcategories. These rapid coding systems make restrictive assumptions about what aspects of interaction are of interest (again focusing primarily on positive vs. negative communication), which can represent an advantage or limitation depending on one's point of view.

Mutually exclusive and exhaustive coding schemes pose conceptual as well as practical challenges. Mutual exclusivity requires the assignment of a single code per unit, although, in theory, messages perform multiple functions simultaneously (Jacobs, 2002; Robinson, 2011). For example, friendly joking during conflict might show affection at the same time that it conveys tacit criticism. Thus, coders must judge the *primary* function of a message relative to the purpose of the coding system. To assist coders, categorical coding sometimes invokes rules of precedence that assign a coding unit to one particular category when it potentially fits multiple categories. For example, the MICS-IV and VTCS (Box 12.1) assign priority to codes seen as more important or as offering clearer interpretation.

Folger et al. (1984) advised against strict adherence to mutual exclusivity and suggested that validity concerns can require one to code each unit into multiple categories or along more than one dimension. However, one can readily see practical limitations to such advice. Allowing multiple codes increases the complexity of coding and subsequent analysis: One must determine when and how to assign multiple codes without compromising reliability, how to collate variable codes per unit, and how to analyze sequential structure if there are multiple antecedent and consequent acts. Instead of multiple codes, another way to address multifunctionality is to use more than one coding system. For example, the CISS has separate codes for verbal content and nonverbal affect. Of course, this approach also multiplies the time and expense of coding.

The conventional requirement of exhaustiveness raises a different conceptual issue. To ensure exhaustiveness, categorical systems routinely include a default category, such as *uncodable*, *other*, or *neutral*, that provides designations for units that are not otherwise classified by the system. Krippendorff (2011) advised against overly broad application of the default category, as this suggests that the coding system is logically incomplete and yields unusable information. An overly broad default category also provides coders with an easy way of avoiding difficult decisions that can be a source of unreliability (Krippendorff, 2011). On the other hand, coding every unit risks overinterpreting messages that lack clear meaning on the dimensions coded. An alternative involves *sieve* coding (Guetzkow, 1950), whereby researchers designate only certain units for coding on the basis of their research aims (Folger et al., 1984). McNulty and Russell's (2010) selective coding of negative messages illustrates this strategy, as does coding of question sequences in physician-patient interviews (Robinson, 2011).

Checklists. When using checklists, coders identify all categories that apply to the coding unit in binary fashion (i.e., each code is either present or absent). Checklist coding methods are especially common in observational studies of parent-child interaction (e.g., Roggman, Cook, Innocenti, Norman, & Christiansen, 2013). The RCISS illustrates the use of a checklist system for coding couple conflict (Krokoff et al., 1989). Checklists might apply to short units, such as speaking turns (as in the RCISS), longer time-based intervals (e.g., Vivian,

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Langhinrichsen-Rohling, & Heyman, 2004), or entire interactions. In contrast to categorical systems, checklist codes are not mutually exclusive and are not necessarily exhaustive. For example, one could code for verbal confrontation without discerning any relevant forms in a given interaction. Checklists thereby simplify coding relative to categorical systems because coders do not have to fit each unit into one and only one category. This makes it practical in some cases to conduct coding "live" during naturalistic observation or to code recorded interactions without transcripts. However, the relative efficiency of checklists can partly rest on application of a relaxed reliability standard, in which reliability is assessed in terms of summary scores (e.g., overall positivity or negativity) rather than unit-by-unit coder agreement (e.g., Krokoff et al., 1989).

Rating Systems

Rating systems involve coders rating the degree to which people display targeted communicative acts. As with the rapid versions of the categorical systems described above (RMICS and RCISS), rating systems typically focus on higher order categories that categorical micro-codes are often combined into. Rather than distinguishing a large list of distinct acts, coders consider a range of relevant acts to determine the presence of broadly defined dimensions, such as positive, negative, and avoidance (Gill, Christensen, & Fincham, 1999; Julien, Markman, & Lindahl, 1989). Researchers using this approach recognize that theoretically relevant dimensions often represent clusters of interrelated acts. These clusters of interrelated acts might not all be exhibited or enacted to the same degree by a particular person. Whereas categorical codes indicate whether a code happens or not, ratings often integrate information on frequency, intensity, and duration to index the magnitude of the targeted act (Margolin et al., 1998).

A good example of a rating system is the Conflict Rating System (CRS; see Box 12.2), which was designed to assess demand-withdraw patterns in couple conflict. Observers watch the entire interaction and rate the degree to which each partner exhibited each dimension (e.g., *discussion*, *blames*, *pressures for change*) during the interaction (1 = *none*, 9 = *a lot*). Coders are instructed to consider the frequency, intensity, and duration of the verbal and nonverbal behaviors relevant to each dimension and to make a judgment of magnitude relative to other individuals in similar interactions. Christensen, Heavey, and colleagues decided to use global ratings to focus on interaction patterns that can manifest in a variety of ways and to assess the intensity rather than frequency of such patterns (Sevier et al., 2004). The resulting ratings distinguish between mild and severe forms of demand-withdraw that may or may not occur at the same frequency. For example, mild but frequent hesitation to discuss topics would produce a lower "withdraws" rating (see codes in Box 12.2) than extreme disengagement and silence that occurred for a shorter time. Balancing frequency with intensity in ratings of magnitude is important because instances of extreme disengagement at pivotal moments in the interaction are likely to have a more pronounced impact on problem resolution and subsequent relationship outcomes (see Sevier et al., 2004).

A central benefit of rating systems is that they reduce the time and expense required to obtain analyzable data while producing similar results as categorical codes (Gill et al., 1999; Julien et al., 1989). Gill et al. (1999) coded couples' conflict interactions using the VTCS (Box 12.1), a categorical code system, and the revised CRS to contrast the utility of each system. The VTCS required more training for coders to reliably distinguish specific codes (about 15 hours) and additional hours to transcribe, unitize, and code interactions. In contrast, the CRS assumes that coders are already equipped with a general understanding of coding constructs and thus require only a short training period to fine-tune this existing knowledge (about 8 hours). Rating entire interactions (vs. speaking turns) directly from video recordings (vs. transcripts and video for the VTCS) took less than an hour per couple. After combining VTCS discrete codes into similar dimensions as the CRS, the scores derived from each coding system were associated. The systems also predicted concurrent and longitudinal satisfaction in similar ways. The one difference, however, was that global ratings of avoidance in the CRS appeared to capture a broader array of communicative acts than those assessed by the VTCS, which could enhance predictive utility but might also reduce understanding of the meaning and impact of specific acts.

Although ratings are an efficient approach to coding, this can be partially offset by the need for multiple raters per interaction to ensure adequate reliability. For example, Gill et al. (1989) had eight raters (four per spouse) analyze each interaction, with reliability based on combined ratings (the Spearman-Brown formula). A single coder applied the VTCS, except for 20% of interactions that were double-coded to check reliability (κ).

Critically, rating systems allow messages to own multiple functions. As described above, in most

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

categorical code systems, observers need to assign one code to each unit, which can involve tough decisions regarding the principal function of the unit. In rating systems, communication can be indexed as a blend of different acts, with the final ratings capturing the relative weight of applicable categories. For example, the CSCS (Box 12.2) organizes ratings into higher order categories that reflect the valence and directness of communication strategies. Partners' communication across the interaction or within a specific speaking turn can be a blend of all four types. For example, a person might try to reason with his or her partner (positive-direct) while also threatening negative consequences if his or her solution is not adopted (negative-direct). The resulting ratings represent the relative magnitude of each type, such as high levels of positive-direct (5 out of 7) and relatively mild negative-direct (3 out of 7) or vice versa. By assessing the relative presence of different strategies, this approach does not truncate assessment to the primary strategy only but still maintains the ability to hone which aspects of communication are most predictive of outcomes. For example, accounting for the associations across direct strategies, Overall et al. (2009) found that both positive-direct and negative-direct strategies were independently associated with greater problem resolution over time. Rating the magnitude of all strategies also avoids the difficulty of trying to classify polysemous (i.e., multiple-meaning) messages into discrete codes.

Rating systems also contain important drawbacks. Global ratings lack detail regarding the specific acts present and therefore which acts might have the strongest explanatory power. Rating systems also lack information about time and sequential contingencies across partners, such as the likelihood that demand prompts withdraw. Although the CRS ratings of one partner's demand and the other partner's withdraw can be combined to create demand-withdraw composites, such an index does not reveal whether withdraw was contingent on (i.e., was influenced by) the partner's demand (Sevier, et al., 2004).

Alternatively, the presence of specific sequences can be rated, such as the degree to which a parent demands and child withdraws across an interaction (e.g., Caughlin & Ramey, 2005). This approach does not constrain assessment of sequences to each turn or unit of analysis (as does sequential analyses). Such lack of constraint proves useful if important interaction patterns occur across wider time spans and, more important, if the time course of dyadic patterns or the length of interaction varies across the sample. In addition, rather than rating the entire interaction, the interaction can be divided into shorter time intervals, rating systems applied to each interval, and then time-series analyses used to test contingency-based predictions. For example, Overall, Simpson, and Struthers (2013) used the CSCS to rate interactions every 30 seconds to test whether positive-indirect strategies by one partner were associated with reductions in withdrawal in the next 30-second interval.

The most important limitation of rating systems might be that they rely heavily on coders' interpretation of the communication exhibited, even more than typical categorical systems. By coding more global categories, rating systems focus on what the researcher believes is theoretically relevant. This helps ensure that the design tests research questions of interest and is valuable when the wider context of the interaction alters the meaning of the same specific act, such as whether advice on how to tackle a problem represents reasoning or autocracy (CSCS; Box 12.2). However, focusing on broader categories asks coders to make inferences about the meaning of observed communication and then aggregate these inferences with frequency and intensity to generate a holistic rating (Margolin et al., 1998). Both the CRS and the CSCS (Box 12.2) adopt a "cultural informant" approach (Gottman & Levenson, 1986), which assumes that coders possess a deep understanding of social interactions, make such interpretations in their day-to-day lives, and thus can reliably decode the meaning of communication. Nonetheless, relying on coders' interpretations inevitably provides more room for idiosyncratic views to bias ratings. In contrast, the descriptiveness of many categorical codes reduces the level of inference required, which may reduce coder bias. Next, we discuss coder inference and bias in more detail.

The Role of Inference in Communication Coding

Sources and Levels of Inference

Although inference is inherent to observational coding (Krippendorff, 2004), it is not always clear what kinds of inferences are carried by communication codes (Folger et al., 1984). Much of the time, observational codes are simply called "communication behaviors," suggesting that codes reference outward features of communication only (i.e., what people "actually" do). Although actual behavior is the starting point for observational research, coding schemes typically do not describe behavior so much

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

as produce structured inferences about functional properties of communication (e.g., messages as forms of affection, social support, or conflict avoidance).

As Stone, Tai-Seale, Stults, Luiz, and Frankel (2012) observed, inferences made by coders can be ambiguous in ways that are not obvious from the usual description of coding procedures. These authors coded illness-related emotions expressed by patients and empathic responses by physicians, phenomena that have parallels in the way couples express and respond to emotionally laden disclosures during conflict. Although they used a previously validated coding system, Stone et al. (2012) found that patient verbal expression of emotion was ambiguous in unanticipated ways. For example, emotion words and other cues were often "fuzzy" and varied from one patient to another; moreover, discussion of illness appeared emotionally laden to coders even in the absence of emotion cues recognized by the coding system.

Coding systems differ in how they resolve such ambiguities. On one hand, a system might restrict attention to readily observable emotion cues, as in automated analysis of affect on the basis of word valence (Baek, Cappella, & Bindman, 2011), facial expressions (Cohn & Sayette, 2010), or acoustic features of speech (Black et al., 2013). Alternatively, coders might identify emotions from context, on the basis of their own implicit cultural knowledge and experience.

The different approaches reflect a distinction between *manifest* (physical or surface) versus *latent* (symbolic) content analysis (e.g., Holsti, 1969). Most obviously, manifest content includes nonverbal behaviors recorded without assistance by human coders or inference about sender intent. Whereas inferences about message intent are essential to interpretation of verbal communication (Jacobs, 2002), Buck and VanLear (2002) argued that many nonverbal behaviors are emitted and apprehended spontaneously (i.e., unintentionally and automatically) on the basis of biologically programed response patterns. Coding of *spontaneous* communication still involves inference, insofar as it rests on theoretical assertions about which manifest cues are important to observe and what functions they serve. Nonetheless, coding of physical cues (e.g., movement of facial muscles) does not require inference about conventional or personal meaning, as does coding of verbal communication or symbolic forms of nonverbal expression.[5] Between strictly manifest and latent content lie forms of coding that involve low-level inferences about speaker intent that are performed easily by any competent language user (e.g., whether a question is rhetorical). However, most interaction coding is more inferential: the codes identify abstract relational events (e.g., confrontation) and associated acts (e.g., criticism). Here again, considerable variation occurs in the discretion afforded to coders. Some systems constrain coder inferences through extensive rules and training, whereas others (such as the rating systems noted earlier) treat coders as cultural informants and allow them greater latitude to fill in meaning.

In addition to the inferences conveyed by coders, a second level of inference occurs when researchers aggregate codes into summary measures. For example, most categorical coding systems confine coder judgments to moderate inferences (e.g., whether an utterance represents acceptance or denial of responsibility) but aggregate on the basis of researcher theories connecting specific codes to summary constructs (e.g., overall positivity or negativity).[6] Notably, coding methods do not always collapse codes in the same way. For example, avoidance and withdrawal are treated as communicative negativity in some systems (RCISS, RMICS) but not others (CRS, VTCS), and problem description may be construed as positivity (RCISS) or neutrality (RMICS). Moreover, researchers often modify constructs ad hoc when collapsing codes. Heyman (2001) noted that researchers have "mixed and matched" codes from the MICS to such an extent that virtually no studies evaluate identical constructs.

Locus of Meaning

Another general principle of message interpretation is that the same overt signals can mean something different to participant versus observer (Surra & Ridley, 1991) or to multiple observers with different frames of reference. Coding methods also assess meaning from varying perspectives. Poole, Folger, and Hewes (1987) identified four such perspectives (see also Poole & Hewes, this volume). *Generalized observer* meanings are those available to any uninvolved onlooker to an interaction (e.g., a vocalized pause), whereas *restricted observer* meanings are derived from application of a specialized interpretive scheme by outsiders (e.g., conversational coherence). *Generalized subject* meanings are available to any member of a cultural or subcultural group (e.g., topic shifts), whereas *restricted subject* meanings are accessible only to relationship insiders (e.g., inside jokes or conflict triggers).

In what domain does most communication coding reside? The perspective of the generalized observer is well represented in interaction research

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

but limited to features that can be assessed through manifest content. Restricted subject meaning is not assessable via observer coding, at least as practiced in quantitative interaction research. Instead, most interaction research spans the boundary of restricted observer and generalized subject meaning. For example, all of the coding schemes in Box 12.1 use specialized interpretive rules applied by trained observers, which suggests restricted observer meaning. However, the systems also rely on coders to use their own cultural knowledge to fill in where coding rules are incomplete; for example, when discriminating friendly from hostile joking or criticism from neutral description on the basis of context.

Herein lies the central dilemma of interaction coding. A primary reason for doing interaction coding is to provide an "objective" (i.e., standardized, outsider) perspective on communication that avoids the biases of self-report data and provides a contrast to participant meaning. However, because it is not always possible to codify interaction constructs in terms of manifest content or clearly identifiable stimulus features, coding methods ultimately rely on intuitive judgments by observers to interpret meaning. An advantage of human coders over automated coding is that coders can use their own cultural knowledge to make sense of implicit features of communication. A limitation is that coders can interject their own knowledge in ways that threaten reliability and validity.

Coder Bias

To the extent that observational methods rely on coders to fill in meaning from cultural knowledge, the methods assume that coders represent cultural or subcultural groups in which these meanings reside. Coding methods also assume that coders can apply cultural knowledge to the specific context under investigation. Coders are usually undergraduate or graduate college students. Students can represent broader cultural meanings when these meanings are widely shared. This should be the case with low-level inferences about speech acts but not necessarily so with abstract relational events. Moreover, student coders often fail to represent the cultural and socioeconomic mix of the sample, which potentially affects interpretation of the acts coded. The relative homogeneity, and therefore interpretation, across student coders might also mean that potentially distinct interpretations are not revealed by reliability checks. Their life and relationship experiences can also mean that student coders are ill equipped with contextual knowledge central to the domain of investigation, such as examining communication during the transition to parenthood, within parent-child dynamics, or in distressed samples, such as people suffering from depression, coping with chronic illness, or facing high levels of violence.

Indeed, as Margolin et al. (1998) noted, life experience, gender, and ethnicity can all affect coder judgments. Male coders have a greater propensity than women to view adult behavior as angry and resentful (Davidson et al., 1996) and to see aggressive behavior in children's interactions (Pellegrini et al., 2011). Gender stereotypes are also likely to affect the way women and men are coded, including the inferred intent behind similar behaviors (e.g., silence as sullen guilt induction vs. withdrawal). Similarly, stereotypes of ethnic and cultural groups can bias coding (Bente, Senokozlieva, Pennig, Al-Issa, & Fischer, 2008). Cultural differences can also affect coder inferences because of the way targeted constructs manifest across cultural groups. For example, cultural differences in the appropriateness of direct conflict (Sillars & Canary, 2013) could mean that interactions that appear contentious or avoidant to observers are not experienced in the same way by cultural insiders.

Coders' own relationship experiences are also likely to affect how they evaluate and infer meaning from other people's communication. The relationship field is replete with examples of individual and contextual factors that shape how relationship events are construed and responded to, such as attachment insecurity, relational standards, or levels of relationship satisfaction. Examining families within diagnostic contexts, such as discussing areas of conflict or supporting each other, will undoubtedly activate associated expectations, preferences, and perceptual sets that affect the way interactions are perceived. People are also highly motivated to maintain positive evaluations of their own relationships, and one way this is managed is by downplaying the positivity of other relationships (e.g., Rusbult, Van Lange, Wildschut, Yovetich, & Verette, 2000). This bias might produce a tendency to perceive others' communication as less constructive or loving than is justified (Gagné & Lydon, 2004). Finally, coders might generate their own understanding of the goals of the research (Harris & Lahey, 1982). By extension, individual coders possess their own conceptions about what constitutes "good" or "bad" communication. Coders' application of these tendencies can potentially undermine the assessment intended by the researcher.

What can be done to counteract coder bias? Margolin et al. (1998) recommended ensuring that coding teams are diverse in gender, culture, and

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

general background, including replacing or combining student coders with coders sourced from the wider community. However, achieving representativeness among coders in relation to the target population may not be practical, and it can lead to other problems, such as the coding schedule's being applied in unintended ways and increasing training time. Nonetheless, coder bias is a significant issue. The potential for bias does not render observational coding invalid or useless; however, we do think it necessary to assess results of coding in light of the limitations of human judgment and the perspectives and dispositions coders bring to the task. Moreover, researchers should take every step to minimize coder bias by structuring, limiting, and monitoring coder inference during the coding process.

Managing the Coding Process

Ultimately, coding procedures are designed to coordinate inferences while maintaining the integrity of coding constructs, which equates to the topics of reliability and validity. Whereas a subsequent chapter provides a comprehensive discussion of reliability and validity (Poole & Hewes, this volume), we highlight how reliability and validity are affected by coding procedures and coder characteristics. Reliability and validity are analogous to the problem of intersubjectivity that is the crux of symbolic communication. To coordinate inferences, coders must apply coding rules consistently and fill in meaning by adopting the perspective of others who operate within a particular (generalized or restricted) meaning domain. The success of this enterprise is affected by characteristics of the coding scheme, coding procedures, and coders.

With respect to the coding scheme, more inferential codes are potentially subject to greater bias, as noted above. More inferential codes also tend to be, but are not inevitably, less reliable. As Krippendorff (2004, p. 20) noted, coders can sometimes read between the lines with remarkable consistency. On the other hand, Stone et al. (2012) ultimately limited their coding of emotional expression to the most explicit examples after attempts to code indirect emotional expression proved unreliable. Similar compromises are built into most coding schemes. Researchers often omit subtle and variable features of communication for reliability reasons, no matter how theoretically heuristic these features might be. The complexity of a coding system also affects intercoder reliability. Heyman et al. (2014) advised that coders generally cannot maintain adequate agreement when there are a large number of subtle codes. However, exceptions exist (e.g., Cegala, McClure, Marinelli, & Post, 2000; Sillars et al., 1984).

Procedures can reduce the burden on coders when categorizing or rating a large number of constructs or difficult to judge constructs. For example, in the CSCS, interactions are coded for one category at a time to ensure that coders focus on the particular influence strategy targeted during that wave. Coding in waves reduces cognitive demand; although coders still need to distinguish between multiple strategies, they only need to assess the strategy they are rating in that wave. Applying rating systems to small time intervals, rather than rating multiple dimensions across entire interactions, has the same benefits and may enable coders to more effectively rate and distinguish between multiple codes. These procedures might also reduce the degree to which coders' subjective evaluations can infiltrate the coding process. Furthermore, additional coding waves can minimize the degree to which the tone of the interaction influences coding. For example, using a separate team of coders to index broad dimensions, such as general valence or problem resolution, can provide a way of ensuring that more specific codes are not "infected" by coders general sense of the interaction.

Although more complex coding systems are not inherently less reliable or subject to bias, they might require more detailed coding manuals, greater rule specification, and more extensive training. A coding manual extends the coding scheme by specifying and illustrating coding rules in detail. A more complete coding manual simplifies coding by anticipating and resolving areas of confusion. Inexperienced coders may expect the coding manual to remove all ambiguity; that is, they assume that there is always a "correct" code under the coding rules. Inevitably, however, examples emerge that the author(s) of the coding manual had not anticipated. Furthermore, even familiar examples can become ambiguous because of a shift in context. In such cases, some unreliability is preferable to perfect reliability achieved through arbitrary decision rules that sacrifice validity. Ideally, observers should code clear examples with a very high degree of consistency and make ambiguous judgments with reasonable (at least above chance) reliability while retaining the spirit of coding distinctions.

The coding manual alone cannot always convey subtle distinctions and ambiguities that must be understood to code reliably. Much of this information is transmitted during the training phase. Even systems that rely on coders' existing culturally relevant knowledge need to organize that knowledge into the constructs and language of the

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

coding system and ensure that coders apply that knowledge in the same way. Coder training typically occurs in a stacked fashion. Coders first get familiar with the manual, and then examples of specific codes and difficult distinctions are used to enhance understanding. For rating systems, examples of levels (e.g., low, medium, high) should also be presented to anchor coders' ratings of relative magnitude. Practice sessions are then conducted, which are used to check coder application, isolate areas of confusion, and build coder confidence. Extensive discussion throughout this process can help identify and clarify any problematic areas and to revise coding rules if needed. Low reliability in this phase provides important information about needed refinements and can assist the researcher in clarifying distinctions, both procedurally and theoretically (see Poole & Hewes, this volume).

The amount of coder training and practice needed is relative to the demands of the coding system. Some codes can be applied reliably by observers after only minimal training. Lorber (2006) had minimally trained raters assess *overreactive discipline* of mothers after receiving a 10-minute introduction to coding. Compared with "gold standard" raters, who participated in weekly training and practice sessions over 8 weeks, minimally trained raters were less reliable, but primarily in terms of mean ratings. Rank order was relatively consistent between coders (r = .61). Furthermore, minimally trained raters had good concurrent validity with raters who underwent gold-standard training (r = .72). These results suggest that minimal training may suffice for assessing relative (vs. absolute) scores for interaction, which is often all that is needed to test hypotheses. However, minimal training is most likely to suffice if coding is confined to surface features of communication (e.g., overreactive discipline was partly defined in terms of yelling, pushing, pulling) and simple constructs that tap shared meanings and experiences among coders (e.g., similar experiences of student coders with parental overreaction).

If two or more coders are reliable, this does not necessarily mean that they applied the coding scheme in the same way any other set of coders would or as the researcher intends. For example, under pressure to improve reliability, coders may independently or collectively improvise ad hoc rules that simplify judgments but transform the meaning of codes (Harris & Lahey, 1982). As much as possible, ad hoc rules should be self-consciously identified and, if appropriate, formalized and incorporated into the coding manual. In that way, one can assess whether coder improvisations maintain the integrity of conceptual distinctions. A common temptation is to fashion an ad hoc default category (i.e., "when in doubt, assign code X") for ambiguous examples. This tendency makes the code less descriptive and offers a potential source of spurious observation, especially when coders apply ad hoc rules inconsistently (e.g., ambiguous examples are interpreted as verbal aggression when the interaction "feels" tense but are seen as neutral communication otherwise).

Coder training typically should not stop after coding has begun. Instead, regular meetings with coding teams provide the opportunity for continual discussion and reflection regarding areas of uncertainty. Reliability problems and discrepancies in codes should be carefully examined as a team to reiterate or refine coding categories and rules. In this way, and throughout the coding process, the researcher explicitly and implicitly clarifies the coding terms. Frequent meetings with discussion of discrepancies help counteract against coders drifting from the coding system. The more interactions that are viewed and coded, the more opportunity coders have to generate their own rules and for idiosyncratic biases to creep into coders' understanding and application of the coding system. Thus, continuous monitoring of reliability and frequent discrepancy discussions are essential to maintaining reliability.

Furthermore, when coders are aware that their ratings are checked, they are more likely to stay on task (Harris & Lahey, 1982). Regular checks also provide the chance to consider the presence of coder biases. Discussing bias openly can help coders recognize the filters they bring to the coding process and, in turn, may reduce the impact coder bias has on the resulting data. However, regular meetings and joint coding also has the potential to produce new rules and definitions, or to create "consensual drift" away from the original meaning of particular categories, as coders' discussions generate shared implicit rules for evaluating interactions (Harris & Lahey, 1982). This drift from the original coding manual may result, as described above, in greater reliability across coders but codes that do not represent the theoretical construct as originally conceptualized. Guidance by a principal assessor to keep coders true to the coding system and to record systematic alterations or formal clarifications may be crucial to prevent this from occurring. However, the assessor must also be reflexive enough to enable coders to query and challenge in order to prevent coders from simply mimicking the investigator's view. Investigators also should ensure

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

they do not label, discuss, or interpret codes in ways that convey the central hypotheses to coders, thereby compromising coder neutrality (Harris & Lahey, 1982). Another way to check consensual drift, and reduce the variability that might occur as coders become more accurate across the sample, is to recode the first 10% to 20% of interactions.

Along with characteristics of the coding system and coding process, characteristics of coders affect reliability and validity. The sources of coder bias noted above highlight that coder demographics can affect the results of coding. Moreover, reliability tends to reflect the similarity of coders in terms of their cultural, educational, and professional background, as well as experience with texts (Krippendorff, 2004, p. 128). College students are the default choice as coders, both for convenience and familiarity with coding constructs. Many of the coding schemes used in clinical psychology and family studies (see Kerig & Lindahl, 2001) require coders with advanced, specialized education (reflecting a restricted observer perspective). However, researchers using systems that rely on lay concepts (generalized subject meaning) could prefer coders without specialized training, because they are less prone to overinterpret interactions. As with decisions regarding the type of coding system used, coders should also be selected according to the aims of the research, the coding being conducted, and the nature of the sample assessed.

Conclusion: Coordinating Perspectives on Communication

Observational coding of communication represents a form of message interpretation that parallels everyday communication but with a formal structure for interpretation and self-reflexive attention to the reliability and validity of inference. As we have noted, most communication coding represents a standardized observer perspective, which combines elements of restricted (theory-driven) and generalized (culturally derived) observer meaning. Observational coding provides an "objective" perspective in the sense that observations are not tainted by involvement in the communication episode and are replicable across observers. A key motivation for doing observational coding is to provide a more objective assessment of communication than participants' own self-reports typically provide. Participant accounts of communication are subject to many known biases, and we often assume that people may not know, or cannot accurately assess, the acts they and others enact during interactions.

Nonetheless, as we have discussed, coding constitutes an inferential act that often reflects bias. Whereas participant perspectives are biased by involvement in communication and other limitations of informal observation, observers are biased by their own goals and experiences. Observers also lack access to insider context that informs meaning for participants, such as relationship history and culture. Thus, we caution against treating observational coding as an unfiltered behavioral description and the only valid or true representation of actual communication. Kerig (2001) summed this point nicely:

> People behave in ways that are discrepant from their self-perceptions, and only direct observation can capture their behavior independently of their appraisals of it. . . . However, saying that the observer has a unique viewpoint does not mean that it necessarily is the most valid one. Observational methods are no more purely "objective" than any other tool in the researcher's toolbox. Underlying every coding category lie choices, and every choice . . . is informed by the investigator's conceptual framework. (p. 2)

In sum, the coding methods we considered in this chapter offer an important way in which social interaction can be assessed. Nonetheless, the value and utility of the outsider perspective must be considered in light of the ways coding methods are applied and, in turn, the degree to which coding procedures rely on or reduce coder inference and bias. We see observational methods as a valuable addition to insider perspectives rather than a superior assessment of communication. Some interaction constructs are best assessed by insider perspectives. Participants' subjective emotional experiences, internal dialogue, and communication intentions are difficult (and perhaps impossible) to discern accurately because insiders' shared histories and understandings influence the meaning of communicative acts (restricted subject meaning). Moreover, regardless of the veracity of people's reports, subjective experiences and perceptions have a powerful impact on people's relationship evaluations and ultimately the courses of their relationships. The most complete approach, therefore, is to assess both insider and outsider perspectives in order to examine how both participants' subjective perceptions and the observable patterns that stimulate and result from participant sense-making shape relationships and the people in them.

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

References

Baek, Y. M., Cappella, J. N., & Bindman, A. (2011). Automating content analysis of open-ended responses: Wordscores and affective intonation. *Communication Methods and Measures*, *5*, 275–296.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Bente, G., Senokozlieva, M., Pennig, S., Al-Issa, A., & Fischer, O. (2008). Deciphering the secret code: A new methodology for the cross-cultural analysis of nonverbal behavior. *Behavior Research Methods*, *40*, 269–277.

Birchler, G. R., Weiss, R. L., & Vincent, J. P. (1975). Multimethod analysis of social reinforcement exchange between maritally distressed and nondistressed spouse and stranger dyads. *Journal of Personality and Social Psychology*, *31*, 349–360.

Black, M. P., Katsamanis, A., Baucom, B. R., Lee, C., Lammert, A. C., Christensen, A., . . . Naravanan, S. S. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication*, *55*, 1–21.

Buck, R., & VanLear, C. A. (2002). Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication*, *52*, 522–541.

Caughlin, J. P., & Ramey, M. E. (2005). The demand/withdraw pattern of communication in parent-adolescent dyads. *Personal Relationships*, *12*, 339–355.

Cegala, D. J., McClure, L., Marinelli, T. M., & Post, D. M. (2000). The effects of communication skills training on patients' participation during medical interviews. *Patient Education and Counseling*, *41*, 209–222.

Cohn, J. F., & Sayette, M. A. (2010). Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. *Behavior Research Methods*, *42*, 1079–1086.

Davidson, D., MacGregor, M. W., MacLean, D. R., McDermott, N., Farquharson, J., & Chaplin, W. F. (1996). Coder gender and potential for hostility ratings. *Health Psychology*, *15*, 198–302.

Erbert, L. A., & Duck, S. W. (1997). Rethinking satisfaction in personal relationships from a dialectical perspective. In R. J. Sternberg & M. Hojjat (Eds.), *Satisfaction in close relationships* (pp. 190–217). New York: Guilford.

Folger, J. P., Hewes, D. E., & Poole, M. S. (1984). Coding social interaction. In. B. Dervin & M. Voight (Eds.), *Progress in communication sciences* (Vol. 4, pp. 115–161). Norwood, NJ: Ablex.

Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review*, *8*, 322–338.

Gill, D. S., Christensen, A., & Fincham, F. D. (1999). Predicting marital satisfaction from behavior: Do all roads really lead to Rome? *Personal Relationships*, *6*, 369–387.

Gottman , J. M. (1979). *Marital interactions: Experimental investigations*. New York: Academic Press.

Gottman, J. M. (1982). Temporal form: Toward a new language for describing relationships. *Journal of Marriage and the Family*, *44*, 943–962.

Gottman, J. M. (1994). *What predicts divorce? The relationship between marital processes and marital outcomes*. Hillsdale, NJ: Lawrence Erlbaum.

Gottman, J. M., & Levenson, R. W. (1986). Assessing the role of emotion in marriage. *Behavioral Assessment*, *8*, 31–48.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology*, *6*, 47–58.

Hahlweg, K. (2004). Kategoriensystem für Partnerschaftliche Interaktion (KPI): Interactional coding system (ICS). In P. K. Kerig & D. H. Baucom (Eds.), *Couple observational coding systems* (pp. 122–142). Mahwah, NJ: Lawrence Erlbaum.

Harris, F. C., & Lahey, B. B. (1982). Recording system bias in direct observational methodology. *Clinical Psychology Review*, *2*, 539–556.

Heavey, C. L., Layne, C., & Christensen, A. (1993). Gender and conflict structure in marital interaction: A replication and extension. *Journal of Consulting and Clinical Psychology*, *61*, 16–27.

Heyman, R. E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment*, *13*, 5–35.

Heyman, R. E. (2004). Rapid Marital Interaction Coding System (RMICS). In P. K. Kerig & D. H. Baucom (Eds.), *Couple observational coding systems* (pp. 67–93). Mahwah, NJ: Lawrence Erlbaum.

Heyman, R. E. Lorber, M. F., Eddy, J. M., & West, T. V. (2014). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 343–370). New York: Cambridge University Press.

Heyman, R. E., Weiss, R. L., & Eddy, J. M. (1995). Marital Interaction Coding System: Revision and empirical evaluation. *Behaviour Research and Therapy*, *33*, 737–746.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Jacobs, S. (2002). Language and interpersonal communication. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 213–239). Thousand Oaks, CA: Sage.

Julien, D., Markman, J. J., & Lindahl, K. M. (1989). A comparison of global and a microanalytic coding system: Implication for future trends in studying interactions. *Behavioral Assessment*, *11*, 81–100.

Kerig, P. K. (2001). Introduction and overview: Conceptual issues in family observational research. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum.

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Kerig, P. K., & Lindahl, K. M. (Eds.) (2001). *Family observational coding systems: Resources for systemic research*. Mahwah, NJ: Lawrence Erlbaum.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures*, *5*, 93–112.

Krokoff, L. J., Gottman, J. M., & Hass, S. D. (1989). Validation of a global rapid couples interaction scoring system. *Behavioral Assessment*, *11*, 65–79.

Lindahl, K. M. (2001). Methodological issues in family observational research. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 23–32). Mahwah, NJ: Lawrence Erlbaum.

Lorber, M. F. (2006). Can minimally trained observers provide valid global ratings? *Journal of Family Psychology*, *20*, 335–338.

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review*, *1*, 195–213.

McNulty, J. K., & Russell, V. M. (2010). When "negative" behaviors are positive: A contextual analysis of the long-term effects of problem-solving behaviors on changes in relationship satisfaction. *Journal of Personality and Social Psychology*, *98*, 587–604.

Notarius, C. I, Markman, H. J., & Gottman, J. M. (1983). Couples Interaction Scoring System: Clinical applications. In E. E. Filsinger (Ed.), *Marriage and family assessment* (p. 117–151). Beverly Hills, CA: Sage.

Overall, N. C., Fletcher, G.J.O., Simpson, J. A., & Sibley, C. G. (2009). Regulating partners in intimate relationships: The costs and benefits of different communication strategies. *Journal of Personality and Social Psychology*, *96*, 620–639.

Overall, N. C., Simpson, J. A., & Struthers, H. (2013). Buffering attachment-related avoidance: Softening emotional and behavioral defenses during conflict discussions. *Journal of Personality and Social Psychology*, *105*, 854–871.

Pellegrini, A. D., Bohn-Gettler, C. M., Dupuis, D., Hickey, M., Roseth, C., & Solberg, D. (2011). An empirical examination of sex differences in scoring preschool children's aggression. *Journal of Experimental Child Psychology*, *109*, 232–238.

Poole, M. S., Folger, J. P., & Hewes, D. E. (l987). Methods of interaction analysis. In G. R. Miller & M. Roloff (Eds.), *Interpersonal processes: New directions in communication research* (pp. 220–256). Beverly Hills, CA: Sage.

Putnam, L. L., & Jones, T. S. (1982). Reciprocity in negotiations: An analysis of bargaining interaction. *Communication Monographs*, *49*, 171–191.

Robinson, J. D. (2011). Conversation analysis and health communication. In T. L. Thompson, A., R. Parrott, & J. F. Nussbaum (Eds.), *The Routledge handbook of health communication* (2nd ed., pp. 501–518). New York: Routledge.

Roggman, L. A., Cook, G. A., Innocenti, M. S., Norman, V. J., & Christiansen, K. (2013). Parenting Interactions With Children: Checklist of Observations Linked to Outcomes (PICCOLO) in diverse ethnic groups. *Infant Mental Health Journal*, *34*, 290–306.

Rusbult, C. E., Van Lange, P.A.M., Wildschut, T., Yovetich, N. A., & Verette, J. (2000). Perceived superiority in close relationships: Why it exists and persists. *Journal of Personality and Social Psychology*, *79*, 521–545.

Sevier, M., Simpson, L. E., & Christensen, A. (2004). Observational coding of demand-withdraw interactions in couples. In P. K. Kerig & D. H. Baucom (Eds.), *Couple observational coding systems* (pp. 159–172). Mahwah, NJ: Lawrence Erlbaum.

Sillars, A., & Canary, D. J. (2013). Conflict and relational quality in families. In A. L. Vangelisti (Ed.), *Routledge handbook of family communication* (2nd ed., pp. 338–357). New York: Routledge.

Sillars, A. L. (1986). *Procedures for coding interpersonal conflict: The Verbal Tactics Coding Scheme (VTCS)*. Unpublished coding manual, University of Montana, Missoula.

Sillars, A. L., Pike, G. R., Jones, T. S., & Murphy, M. A. (1984). Communication and understanding in marriage. *Human Communication Research*, *3*, 317–350.

Sillars, A., & Vangelisti, A. L. (2006). Communication: Basic properties and their relevance to relationship research. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge handbook of personal relationships* (pp. 331–351). New York: Cambridge University Press.

Sillars, A. L., & Wilmot, W. W. (1994). Communication strategies in conflict and mediation. In J. Wiemann & J. A. Daly (Eds.), *Strategic interpersonal communication* (pp. 163–190). Hillsdale, NJ: Lawrence Erlbaum.

Stone, A. L., Tai-Seale, M., Stults, C. D., Luiz, J. M., & Frankel, R. M. (2012). Three types of ambiguity in coding empathic interactions in primary care visits: Implications for research and practice. *Patient Education and Counseling*, *89*, 63–68.

Street, R. L., & Cappella, J. N. (1985). Sequence and pattern in communicative behavior: A model and commentary. In R. L. Street, Jr., & J. N. Cappella (Eds.), *Sequence and pattern in communicative behavior* (pp. 243–276). London: Edward Arnold.

Surra, C. A., & Ridley, C. A. (1991). Multiple perspectives on interaction: Participants, peers, and observers. In B. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 35–55). New York: Guilford.

Vivian, D., Langhinrichsen-Rohling, J., & Heymay, R. E. (2004). The thematic coding of dyadic interactions: Observing the context of couple conflict. In P. K. Kerig & D. H. Baucom (Eds.), *Couple observational*

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

coding systems (pp. 273–288). Mahwah, NJ: Lawrence Erlbaum.

Watzlawick, P., Beavin, J., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. New York: Norton.

Endnotes

1. The original version of the VTCS collapsed into three macro-categories (i.e., *integrative*, *distributive*, and *avoidance*) but was revised to reflect more descriptive macro-categories that avoid a priori assumptions about which messages serve positive or negative functions.

2. An even more recent system that evolved from the CRS, the Couples Interaction Rating System, has summary scores for demand and withdrawal but lacks the positive and negative scales of the CRS (see Sevier, Simpson, & Christensen, 2004).

3. A thought unit is a segment of speech that expresses a single, unified thought.

4. Although CA and interaction analysis are separate research traditions with very different methods and assumptions, Robinson (2011) argued that the two approaches can form a symbiotic relationship. In an observational study of physician-patient interviews, CA insights gleaned from close analysis of individual interactions have informed development of traditional coding schemes, thereby contributing to validity. Traditional coding methods have helped demonstrate that CA informed distinctions matter by documenting their statistical association with outcomes (Robinson, 2011).

5. In practice, it can be nearly impossible to discern the difference between spontaneous communication and intentional manipulation of the same signals (i.e., *pseudo-spontaneous* communication; Buck & VanLear, 2002). Regardless of their true origins, nonverbal signals may be interpreted at the level of manifest content or symbolic meaning.

6. The same may be said for coding of relational control (see Rogers & Cummings, this volume), which begins with low-level inferences about the grammatical and pragmatic form of utterances but aggregates specific codes into patterns of dominance and domineeringness.

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.