10 Using direct observational methods to study the real lives of families

Advantages, complexities, and conceptual and practical considerations

Rena L. Repetti, Shu-wen Wang and Meredith S. Sears

Whether describing jobs, families, or work–family processes, research on the interface between work experiences and family life has relied almost exclusively on participant-report measures in which the respondent describes the self or another target, such as a spouse. These methods offer an obvious advantage to researchers in that they are relatively inexpensive and easy to use. But there are shortcomings, and direct observational methods offer an alternative. This chapter focuses on the use of direct observational methods to address research questions in the work–family field, particularly their use in the assessment of families. We begin by discussing sources of error in participant reports and then move to a discussion of key issues involved in designing a direct observational study and in making sense of the data. Throughout, we distinguish between naturalistic designs and approaches that incorporate greater researcher control over what is observed.

Measuring work and family constructs through indirect and direct observations

Psychological constructs are explanatory variables that are not directly measurable but make sense of the complexity of human behavior. Constructs such as interpersonal conflict, positive and negative emotion, social support, spillover, and cross-over are common and meaningful in the work–family research literature. Instead of directly observing these phenomena, indicators or manifestations of them are observed and recorded, either by researchers or the participants themselves. The approaches used to assess manifestations of a construct rely on some form of either indirect or direct observation. With indirect methods, such as questionnaires, the research participant describes his or her observations of a target (e.g., experiences on the job, a family characteristic, a spouse's behavior). With direct observational techniques, the work or family variable is observed in the here-and-now by a third party. Direct observations have been used to assess both family interaction (Costigan, Cox and Cauce, 2003; Grossman, Pollack and Golding, 1988; Repetti and Wood, 1997; Wang, Repetti and Campos, 2011) and work conditions (Repetti, 1989) in studies addressing work–family questions. As we move from indirect to direct observations, the work of observing the target and

of processing and recording those observations shifts from the participants to the researcher.

Researchers use constructs to organize observations, which can be represented by anything from numbered ratings on questionnaires to images on video. Although all measurement techniques – whether based on participant reports or direct observations made by third parties – involve processes of choosing, defining, and interpreting constructs, the approaches vary with respect to the point at which the construct is defined and the process by which it is interpreted. Because constructs themselves are not directly observable, decisions must be made about the indicators that constitute the construct (e.g., what comprises family conflict) and whether the manifestations of the construct have been observed. The direct and indirect measurement approaches discussed in this chapter vary according to how those decisions are made, and how participants and researchers contribute to the decision-making process.

The phenomena of interest to psychologists (i.e., the manifestations of a psychological construct) can be thought of as roughly falling into two categories. Conscious "internal" psychological processes – such as beliefs, thoughts, and feelings – can only be described by the individual. A larger class of externally measurable, observable phenomena – such as behaviors that constitute marital conflict or emotion expression, job characteristics, or indicators of mental and physical health – can be described by a range of observers. These might include the individual him or herself, a family member or co-worker, or an independent observer, such as a researcher. When the indicators of a construct are subjective experiences and thought patterns, such as participants' *perceptions* of their experiences or the environment, self-report instruments are ideal. For example, marital satisfaction is a consequence of an individual's subjective cognitive and emotional experience of a marriage, and can only be addressed by self-report. Though most constructs in the field of work and family research refer to observable phenomena, they are almost always assessed at the level of individual impressions only. We argue that, when constructs are manifested by phenomena other than purely subjective experiences and perceptions, there is an advantage to supplementing self-reports with observations made by others. The field's understanding of familiar constructs, such as work–family balance, would benefit from integrated methods that assess both internal as well as external phenomena. That's because we do not yet know the extent to which work–family balance is a psychological concept that exists in the minds of individuals, and the extent to which it is enacted in social space shared with family members, co-workers, and supervisors.

Sources of error in participant reports

The limitations of participant-report methods most commonly mentioned in the work–family research literature are the predispositions, attitudes, and beliefs that participants bring to the task of responding to items on questionnaires and surveys. Respondent biases – whether due to social desirability, mood, or another individual

characteristic – introduce systematic error and, as a result, correlations between variables assessed with scales completed by the same person are inflated. For example, cross-sectional correlations between self-report scales assessing job satisfaction and marital satisfaction are sometimes attributed to "positive spillover" between experiences in the two settings. However, it is impossible to distinguish any actual spillover effects (work influencing married life, or vice versa) from individual characteristics that contribute to responses on both scales. When responding to items like "I feel good about my job," or "how happy are you in your marriage," an individual's tendency to look at the glass as "half empty" would lead to a focus on deficits, or what could be better in *both domains*. On the other hand, a wish to view one's life in the best possible light would direct attention to positive experiences *both at work and at home*, or perhaps to comparisons with people who are in worse situations. In either case, the participant's responses would be biased in the same direction on the two scales, and the correlation between them would at least partially reflect that response bias. The consequence of this common method variance is an increase in Type I error, or false–positives in the research literature. In short, a reliance on cross-sectional correlations between self-report measures in the work–family field has undoubtedly led to an overestimation of the robustness of concepts like spillover. To address that problem, rather than rely on self-report data alone, work–family researchers sometimes include observations made by other participants, such as spouses or coworkers (Repetti, 1994; Story and Repetti, 2006; Schulz, Cowan, Cowan and Brennan, 2004).

There are also a number of sources of random error in measures that rely on the reports of participants including differences among respondents in the way that questions and response scales are interpreted. The reliability of a scale provides information about the degree to which individual users respond similarly to items designed to assess the same construct or to the same items on different occasions; it does not provide information about the extent to which respondents differ in their interpretation of key terms or in their use of a response scale. Although the researcher may have one meaning in mind for a term on a scale, participants may introduce nuances in their interpretation of the term. For instance, when responding to a typical item on a family conflict scale, such as "we fight a lot in our family," participants may conceive of the terms "fight" and "a lot" in different ways. "Fight" can be interpreted to mean overt physical aggression, or a heated verbal argument, or more subtle forms of aggression without any overt hostility at all. Because the language used on a scale is open to interpretation, each respondent plays some role in shaping the way that the construct is defined in his or her data. Thus, for example, the strategy of testing correlations between a self-report family conflict scale and a measure of job stress to assess negative spillover is undermined if the interpretations of family conflict and job stress items vary randomly across participants, thereby introducing noise that obscures the identification of true underlying associations. The consequence of non-systematic measurement error is an increase in Type II error, or false–negatives in the research literature. The failure to detect true associations, which is especially likely when effect sizes are

small, is just as consequential for advancements in the work–family field as are false–positive associations due to systematic biases. Aggregating ratings of a single target made by multiple participants is one way to reduce both types of measurement error. For instance Repetti (1987) averaged coworkers' ratings to assess the social climates of bank branches.

Another source of random error is the simple inability to accurately provide the information that is being requested. The description of behavioral and interactional patterns places a substantial cognitive burden on participants with respect to basic recall as well as the aggregation of memories across many contexts and occasions, some of which may be more or less salient depending on the participant's mood or recent events (Stone and Shiffman, 2002). To even attempt to respond with accuracy to the sort of scale items used in work–family studies involves complex processing of information. The example used above ("we fight a lot in our family") requires the respondent to first operationalize terms like "fighting" and "a lot" and then apply those definitions in a consistent manner to daily memories that extend back over a long period of time and countless situations. Corrections must be made to the estimating process. For instance, respondents for whom there is little family conflict to recall must overcome a tendency to overestimate the frequency of fighting from recall of a single (but rare) or recent fight in the family. Of course, measurement error is compounded as the phenomena that participants are asked to describe become more complex. Participants in work–family studies are often asked to describe psychological processes, such as positive and negative spillover and coping responses to work stress. For example, they might be asked whether "stress at work makes you irritable at home." Responding to that item requires – in addition to a shared understanding of what the researcher means by "stress" and "irritability" and a keen awareness and accurate reporting of both – an ability to correctly infer whether irritable behavior, mood, and attitudes expressed at home are *caused by* stress at work. Making the causal inference involves a comparison of periods when work is more and less stressful, and decisions about when behavior at home is a response to job conditions and not some other factor. A complex cognitive calculus is needed to compute changes in behavior and affect and accurately link those changes to specific conditions in the environment. Items like this are just as likely to assess a respondent's *beliefs* as they are to reflect actual associations between experiences and behavior in the work and family domains. However, self-report measures of complex constructs like work–family conflict and spillover are rarely interpreted as measures of impressions, attitudes, or beliefs.

Daily report studies reduce the information processing demands placed on the participant by inquiring about specific behavior or events during a recent limited period of time, such as that day, or the last few hours (Stone and Shiffman, 2002). Clearly less information needs to be recalled and processed, resulting in less error, when responding to a family conflict item on a daily report scale such as, "my husband and I argued tonight," compared to the item, "we fight a lot in our family." The influence of social desirability is likely to be more limited in reports about behavior on a single day compared to asking participants what they or their

families are "usually" like. More important, the processing demands shift from the participant and onto the researcher. For example, in order to determine what participants or their families are "usually" like, person–level or family–level variables can be computed from daily–level data. Moreover, the respondent is relieved of the task of calculating a connection between experiences and behavior at work and at home. That calculus is now performed through statistical analysis of the repeated measures.

Direct observation of families

Direct observation provides a drastically different approach to the definition and examination of constructs than that offered by participant reports. With direct observation, constructs are assessed through phenomena that are directly surveyed by the researcher, bypassing participants as reporters. By gaining a fly-on-the-wall view that directly taps into the behavior or events of interest, these methods allow researchers to more objectively assess what is occurring in real-time rather than rely on subsequent accounts delivered by participants. We believe greater use of direct observational data would benefit the work–family field. Because they remove much of the subjectivity and noise inherent in participant reports, direct observations reduce the amount of Type I and Type II error contained in correlations based purely on self-report scales. This section begins with a discussion of the design and data collection stages of direct observational research, focusing on factors that underlie decisions about whether to observe families in the laboratory or in the field and the issues that arise in choosing specific design features and methodologies for naturalistic studies. Next, we discuss characteristics of observational data, and the steps involved in organizing and coding observations. Throughout, we emphasize naturalistic observational data and the ways in which they differ from observations made in the laboratory.

Designing a direct observational study of families

Methods used by researchers to directly observe families vary with respect to the researcher's role in deciding what about a family or its life is observed, as well as where, when and how the family is observed. The circumstances under which families are observed range from the artificial setting of the research laboratory – in which families might be instructed about what to discuss, for how long, and with whom – to field settings in which families can be recorded as they go about their daily lives. An overriding issue that differentiates among the various approaches to direct observation methods is the degree of control exerted by the researcher.

Choosing the observational situation

Characteristics of the settings in which families are observed include physical location as well as situational factors, such as who is present and the specific

instructions given to family members. Most direct observational assessments of couple or family behavior take place in the structured setting of the laboratory where, typically, a situation is designed to elicit the behavior that constitutes the construct of interest. Researcher demands that shape what occurs in the laboratory constrain participants' behavior such that a restricted but relatively uniform view of a specific process is observed. For example, there is a rich tradition in the marital observation literature of eliciting conflict by asking couples to discuss a highly problematic issue in their relationship for a designated period of time (for a review, see Heyman, 2001). A relatively newer area of marital investigation involves the observation of marital support processes; in these laboratory studies, the researcher structures opportunities for the solicitation and provision of support by having each member of a couple discuss something personal that they would like to change about him or herself (e.g., Pasch and Bradbury, 1998). Behavior in the laboratory may be influenced not only by instructions (e.g., topics for a dyad to discuss, or a particular game for the family to play), but also by space and movement limitations (e.g., having an individual connected to a blood pressure monitor throughout a task), the people present (e.g., a parent–child interaction task that does not include siblings or the other parent), and other controls imposed by the researcher on the observational setting. Attempts to study family behavior in 'seminatural' settings designed to approximate an actual home (e.g., 'apartment laboratory') have revealed that conflict observed in a more typical laboratory setting differed from conflict observed during dinnertime in the 'apartment laboratory' (Gottman and Driver, 2005). Overall, the evidence suggests that behaviors and processes observed in the laboratory do not necessarily map onto the behaviors and processes that unfold in everyday life (Gottman and Krokoff, 1989).

Observations are also made outside of researcher-controlled environments, in locations where family members can be tracked as they go about their normal routines with minimal interference from the researcher, a procedure we refer to here as "naturalistic observation." In comparison to data collected in the artificial setting of most research labs, the ecological validity of observations made in the field is assumed to improve the external validity, and therefore the generalizability, of a study's results. Family interactions have been systematically observed in a variety of natural settings, ranging from the privacy of the family home, to semi-private settings such as daycare centers, to public places like playgrounds or museums. Of course behavior (whether observed by a researcher or not) is also constrained by situational demands in natural settings. For example, emotions and behavior in private settings (e.g., the home) have been found to differ from those displayed in public settings (e.g., the workplace; Larson, Richards and Perry-Jenkins, 1994).

Of course, just because an observation is conducted in the field doesn't necessarily mean that the targets of the observation are behaving in a completely natural manner as they are being observed. First, the level of intrusiveness on the part of the researchers (e.g., how cameras are placed, who is present) may influence how free participants are to act in a "normal" way. Observations made outside of

the laboratory are also sometimes directed in some manner by the researcher. For instance, to study parent–child interaction at home, parents might be asked to play with their infant for 15 minutes as they normally would (e.g., Costigan, Cox and Cauce, 2003), or they might be asked to arrange a family dinner to be filmed (e.g., Hayden *et al*., 1998). There are several reasons why prompted behavior in response to a researcher's instructions, regardless of the setting, would not necessarily map on to everyday behaviors and experiences. Missing are the normal sequencing of events, behaviors, and feelings that usually precede, and give rise to, a particular type of interaction. The presence (or absence) of certain family members can also reduce the observations' external validity. For instance, the researcher might ask that all family members be present for a dinner or that siblings not be present during parent–child play in the home, conditions that may not reflect the social environments in which these events typically occur. Consider, for example, the Costigan *et al*. (2003) study, in which mothers' descriptions of their social climate at work were linked to the observed quality of both parents' interactions with their infant at home; a less supportive job climate was associated with more negative and intrusive, and less positive behavior. The external validity of the 15-minute "play session" was reduced to the degree that the investigators' arrangement of the observational setting did not match the circumstances of everyday life at home.

In addition, just as with participant reports, there are individual differences in social desirability biases that lead to varying degrees of self-correction or editing of behavior. Of course, a research participant has much more control over self-presentation when deciding which number to circle in response to a question about how she argues with her spouse than when asked to discuss a conflict-laden topic with her husband. Not only is her own emotion aroused in the latter situation, her spouse's emotion is also aroused and his behavior is not under her control. Family members in naturalistic observational studies have the least awareness of and control over what they reveal. In addition to having only limited control over the events that occur and the behavior and activities of others as their everyday lives are recorded, they cannot know all of the potential foci of interest in the observations that are collected. Indeed, as discussed below, even a researcher may not be aware of all of the constructs that ultimately will be identified and studied once the data are collected.

Although naturalistic family observations maximize external validity, there are pragmatic considerations for researchers to consider. Extra resources are needed to collect direct observational data in natural settings. Study recruitment can be especially challenging when asking families to allow one or more researchers into their lives. Home observations can be uncomfortable for participants and one uninterested family member can prevent the rest of the family's entry into the study. In addition, a large and well-trained staff is required to collect and process the data. And, if the construct of interest has a relatively low base rate occurrence (e.g., overt family conflict), researchers may have to collect data for quite some time before capturing even a minimal amount of relevant behavior. Low base rates are one of the primary reasons psychologists observe families in controlled

settings in which a desired behavioral situation can be elicited. All of these factors, as well as providing sufficient monetary compensation to participating families, can become quite expensive compared to the typical laboratory-based study. The next two sections focus on occasions when a researcher decides that the advantages of naturalistic observation outweigh the costs and challenges. Two critical design issues are discussed: decisions about the timing of data collection and the choice of recording methods.

Naturalistic methods: Deciding when families are observed

Naturalistic studies can focus on a particular time of day in which an identified interaction or behavior routinely occurs. In a sense, control over the timing of naturalistic observations is comparable to the intentional elicitation of particular social situations in laboratory studies. Naturalistic participant reports such as daily diaries and other momentary techniques (e.g., Ecological Momentary Assessment, Shiffman, Stone and Hufford, 2008; Experience Sampling Method; Larson and Csikszentmihalyi, 1983) describe events as they play out in the natural environment by tracking perceptions and behaviors on repeated occasions. Our understanding of complex phenomena like work–family spillover has been advanced by studies relating daily conditions at work to same-day mood and behavior as assessed by self- and spouse-reports (Story and Repetti, 2006; Schulz, *et al*., 2004). For example, in a sample of air traffic controllers, Repetti (1989, 1994) studied work–family spillover by examining associations between daily reports of job stress and daily descriptions of marital and parent–child interactions later that evening. The data showed same-day linkages between reports of job stress and the level of social involvement and anger in family interactions later that evening. Daily behavior was also linked to direct observational measures of workload: visibility at the airport during the air traffic controller's shift each day (based on hourly weather observations from the National Climatic Data Center) and FAA records of daily air traffic volume. These studies show how repeated measures can be used to assess psychological and behavioral processes as they unfold in real time.

Well-timed direct observations of families in their natural settings can be incorporated into daily research designs. Although the isolation of naturally-occurring events that families or individuals have in common is often a knotty process, "universal" events have been studied through naturalistic designs, particularly when the participants share some common characteristics. For instance, Repetti and Wood (1997) assessed mother–child interactions in a direct observational study of employed mothers as they were reuniting each evening with their preschool-aged children at a worksite daycare center. Behavioral coding of the videotaped reunions indicated that the mothers spoke less and appeared less emotionally involved with their children (e.g., fewer verbal and nonverbal expressions of affection) during reunions that followed more difficult days at work. Thus, an analysis of mothers' daily reports of job conditions in combination with direct observations of their behavior at the moment of reunion with their

children at the end of each workday suggested that daily job stress was often linked to a short-term increase in parental withdrawal. This study allowed researchers to observe a purely naturalistic (i.e., not prescribed) event while maintaining some control over the parameters of the situation.

The best example of the use of naturalistic video data to study work–family processes is found in the work of the UCLA Center on Everyday Lives of Families (CELF). With funding from the Alfred P. Sloan Foundation, our interdisciplinary group studied 32 middle-class dual-earner families with school-age children over the course of a week. Multiple methods, including questionnaires, daily self-reports, interviews, home tours, and salivary cortisol samples were used, but the hallmark of this project is the ethnographic videotaping of everyday family routines and interactions. Family members were filmed as they went about their daily lives inside and outside of the home over two weekdays and two weekend days, permitting direct observation of family behavior and process (Ochs, Graesch, Mittmann, Bradbury and Repetti, 2006). Using the CELF data, Wang, Repetti, and Campos (2011) coded rates of talking, social engagement (i.e., intensity of response to another's initiation of interaction), and negative emotion expression (e.g., frowns) in employed adults with their spouse and children in the first hour after they returned home from work. When those direct observations were linked with self-reported levels of job stress and trait neuroticism (i.e., emotional instability), two pictures emerged. On the one hand, a pattern of negative emotion spillover was observed among husbands and fathers who scored high on neuroticism: more job stress was associated with more involved and negatively valenced interactions. On the other hand, a pattern of social withdrawal was found among the men who scored low on neuroticism: more job stress was associated with less social interaction and less negative emotion expression. There were no associations between job stress and the after-work social behavior of the women in this sample. The contrast with typical questionnaire methods to assess work–family processes is striking. It would be unreasonable to expect research participants to recognize or describe the kind of subtle behavioral changes that were observed in conjunction with job stressors. This approach provides a more detailed and accurate picture of the links between experiences at work and behavior at home. While the larger observation effort recorded hundreds of hours of footage inside the home, this particular investigation "zoomed in" on the first hour window of after-work behavior in order to study work–family process. Thus, timing at the level of coding and analysis – in addition to timing in the initial filming effort – is critical to research that uses naturalistic observations of continuous and non-specific streams of behavior.

Naturalistic methods: Deciding how families are observed

There are a variety of methods that researchers use to document family behavior in natural settings. In addition to video and audio recordings, trained observers can directly register their observations immediately as they occur through notes made on paper or into electronic devices. For example, the CELF study

supplemented video with a scan-sampling method in which live observers used hand-held computers to note the locations and describe the activities in which family members were involved at 10-minute intervals, such as watching TV, preparing dinner, doing homework, talking on the telephone, playing with a video game, or brushing teeth; the activities were later sorted into larger categories such as leisure, housework, communication, and personal care (Ochs, *et al*., 2006; Saxbe, Repetti and Graesch, 2011). The scan-sampling observational data were used to address questions about how activities at home influence physiological recovery after a day at work. Findings indicated that an individual's recovery from the work day – as assessed by cortisol, a stress hormone – was associated with his or her spouse's observed activities at home. Physiological unwinding was enhanced among women whose husbands spent more time involved in housework activities, whereas men's physiological unwinding was more effective if their wives apportioned less of their time to leisure (Saxbe, Repetti and Graesch, 2011). Another recording approach has live observers describe families during set periods of time according to a predetermined reporting scheme. In one study, an observer with a hand-held computer followed family members at home and noted the occurrence of specific behaviors. The variables of interest had already been operationalized; for instance, during the 20-minutes each hour that the target child was observed, behaviors such as hitting, yelling, and destructiveness were noted (Patterson and Forgatch, 1995).

Technological advancements have permitted the recording of behavior and interaction such that continuous streams of raw data are captured prior to any coding or data reduction. Audio-recordings can be taken in the field by the use of devices worn on the person as he or she goes about his or her daily life. One such audio-recording apparatus is the Electronically Activated Recorder (EAR, Mehl, Pennebaker, Crow, Dabbs and Price, 2001), an event sampling device that has been used to sample brief clips of participants' acoustic social environments, capturing everything from self-talk, to laughter, to conversation. For example, Slatcher and Robles (in press) assessed children's social conflicts at home with this methodology. Direct observations can also be collected using video-recordings that capture both auditory and visual information that build archives of rich multisensory data. Videotaping captures non-specific streams of behavior and interactions that occur in a given context. The CELF study is one example of continuous videotaping of family interactions inside families' homes and community settings (e.g., Campos, Graesch, Repetti, Ochs and Bradbury, 2009; Ochs *et al*., 2006; Wang, *et al*., 2011).

Making sense of family observations

Whether in the laboratory or in the field, direct observation affords windows onto behavior and interactional processes for researchers to study in real-time. In order to gain an understanding of families from direct observations, constructs must be operationalized through coding and analysis of the data. The variety of behaviors and situations captured by the observations, the degree of detail contained in the

recordings, the level of order and organization inherent in the raw data, and the uniformity of information across participants have important implications for behavioral coding.

Detail and variability in observational data

The specific features of a direct observational study of families determine the depth and breadth of the data that are collected. A much broader span of behavior is observed in naturalistic studies than when participant behavior is to some degree scripted and controlled by researchers. Naturalistic observations are subject to the idiosyncrasies of the phenomena as they naturally occur. For example, an investigation of dinnertime interactions in the CELF study found that the family composition sitting down to dinner (e.g., in unison with Dad missing, in partial unison with all members present at some point of the dinner), lengths of mealtime (range of 16–40 minutes), types of concurrent activities (e.g., watching TV, doing homework), and how the food was prepared (e.g., cooked from scratch, take-out) all varied considerably (Ochs, Shohet, Campos and Beck, 2010). Conversely, a laboratory investigation of family dinners might invite family members into a structured setting (e.g., an "apartment laboratory"; Driver and Gottman, 2004) in which these factors vary to a lesser extent. In a controlled setting there is more clarity and uniformity in what is captured on film. In fact, the point of researcher control over settings and instructions is precisely that – to reduce variability and limit the scope of what is observed. The goal of live observers' notes is the same; they are designed to limit and organize what is recorded. Both approaches seek to impose order and organization on observations during the data collection stage. The strategy is analogous to indirect observational methods in which items are selected in advance to direct the attention of participants to the particular phenomena (e.g., perceptions, behaviors, experiences) that the researcher wants to assess.

The breadth and level of detail in the raw data that are collected is also determined by the recording methods that are used. For example, video recordings include information about both verbal and nonverbal behavior whereas audio recordings gather verbal behavior only. Setting characteristics influence the level of detail that can be captured in the video or audio recordings. The investigator can collect more detail about certain variables by including the ability to "zoom in" or "zoom out" on the phenomena of interest. For instance, an observational study using videotapes in the laboratory can deliver close-up views of discrete facial movements that reveal miniscule details about facial expressions, a prerequisite for such widely used emotion coding systems as the Specific Affect Coding System (SPAFF; Coan and Gottman, 2007) and Facial Action Coding System (FACS; Ekman, Friesen and Hager, 2002) which require direct unobstructed footage at close proximity. This level of clarity and detail in vantage point is conceded with video observations taken in the field, which contend with shifting angles and movement, ambient noise, changing actors, and spontaneous activity in pursuit of the advantages of ecologically valid data. Similarly, the

amount of detail derived from the work of live observers in naturalistic settings will fluctuate depending on the method by which observations are recorded. For instance, observers who use technologies that facilitate note-keeping (e.g., voice recorders, handheld computers) have the potential to quickly record more detailed observations than if they relied on hand-written notes alone. Additionally, the structure of observers' note-keeping – for example, registering the occurrence of specific pre-determined behaviors or activities versus open-ended descriptions – will also determine the amount of detail and the scope of the data.

Ultimately, researchers who use direct observation are faced with the monumental task of imposing order on the data in order to measure the constructs of interest. Across methods, there are inherent differences in the extent to which the data are collected in an organized manner from the start versus imposing organization *after* the raw data are collected. With participant reports and live observer notes, there is an effort to define key constructs before the data are collected. Similarly, when researchers control family observations through instructions to participants, the constructs are largely defined and decisions made about how to organize observations prior to data collection. For work–family scholars who are considering the use of direct observational methods and are clear about the particular family circumstances and behaviors of interest, it may be optimal to use settings and data recording methods that limit the breadth of behavior and variety of situations that are observed. For instance, in the study mentioned above, reunion behavior after work was assessed by observing employed mothers at a worksite daycare center just as they were reuniting with their preschool-aged children at the end of each work day (Repetti and Wood, 1997). Statistical analyses are more efficient when the design of a study and the collection of data were organized with specific settings and constructs in mind. In the case of naturalistic observations, the process of defining and operationalizing constructs often involves a series of steps *after* the data have been collected in which the raw data are divided into meaningful units, coded in some manner, and analyzed. However, the more that the organization was imposed prior to data collection, the less breadth is present in the data that are collected.

Parsing family observations

With continuous streams of raw observational data, the process of imposing order and meaning begins with the *parsing* of the data prior to the implementation of a coding scheme. In laboratory observation, the constructs under study are largely defined by how they are elicited by the researcher, limiting the degree to which any breaking down of the data is required. For example, there will be fewer components to an interaction that lasts a predetermined and manageable 10 minutes for all families, involves the same actors (e.g., a parent–child dyad), and results from the same scripted prompts delivered by the researcher. Direct naturalistic observations, however, yield uninterrupted streams of behavior because little organization was imposed prior to data collection. For statistical analysis, the raw naturalistic data must be divided into units for analysis; consider the challenge of coding hundreds of

hours of continuous video footage that varies across and within participants in terms of situations, actors, settings, filming angles, and length of time on-screen. The researcher can choose to parse the data in any number of ways. For example, one approach is to organize the observations according to set intervals in which clips of a specific duration are identified. Other parsing approaches may be more content- or process–specific, and thus anchored to *who* appears on-screen (e.g., a specific family member or dyad), and/or *what* is occurring (e.g., routine events, the occurrence of particular types of behavior or activity).

The extra effort involved in parsing data can be illustrated by comparing the observations of parent–child reunions in the worksite daycare study and in the CELF study. In the first case, the observational data were contained in 10-minute videotaped reunion episodes all of which took place in the same room with a table, two chairs, and some toys (Repetti and Wood, 1997). In the CELF study, a parent–child reunion had to first be defined; we examined the two minutes of video recording that began after the parent's initial arrival home and coded the returning parent's first encounter with children who were at home at the time of arrival. Even with this specific definition of a reunion, the CELF data contained many sources of variance that were not represented in the worksite daycare study: the particular family members who were at home each day; even if a child was at home, whether or not the reunion took place within the 2-minute time span; the length of time a child had been at home when the parent arrived; and interruptions, distractions, and activities that took place during the two-minute period, such as phone calls, opening mail, folding laundry, playing on the computer and homework. The added variables in the CELF video, which complicated the identification and coding of reunions, proved to be critical. We learned, for example, that fathers tended to arrive home later than mothers and were therefore more likely to be immediately involved in ongoing household logistics and routines. Fathers were also more likely to be greeted by children who were distracted by an activity, such as homework (Campos *et al*., 2009).

Investigations from the CELF study have used a variety of approaches in the analysis of its large archive of naturalistic observations of everyday family life. For example, the Wang *et al*. (2011) analysis of the first hour of interactions with family members after work adopted a "thin-slicing" approach that organized the video into 10-minute intervals and designated the first 30 second clip at the start of each interval for coding. Other investigations have honed in on specific family routines. For example, in the CELF dinner study described earlier, Ochs and colleagues (2010) identified family dinners – defined as the main household meal of the evening – in order to examine the level of participation of each family member in the meal and the length of the meal. As mentioned above, Campos and colleagues (2009) examined another family routine, daily reunions; in that case, greetings that a parent received from children and spouse were coded when the parent arrived home after work.

In several current projects, CELF researchers are targeting the occurrences of particular types of behaviors or activities, such as supportive interactions in couples, parent–child conversations about peers and academic topics, expressions

of anger in school-age children and parent responses to child negative emotions. When a particular type of situation is the target, the parsing process takes place in multiple steps that progressively impose greater order and organization onto the data. First, clips are identified in which the individual (e.g., school-age child) or dyads (e.g., the couple) of interest appear on screen. Second, occasions are selected when the specific behavior or activity occurs (e.g.s., a child expresses an emotion, an opportunity for a supportive couple interaction occurs, a certain topic is mentioned in conversation). Only then is a detailed coding system applied to rate the phenomena of interest. Note how the first two steps, both part of what we call the parsing process, serve the function that is represented in a laboratory study by researcher control over the social situation in which the family is observed.

The efficiency of the process of searching and parsing video is largely determined by the manner in which the data are organized. A critical early accomplishment at CELF was the organization of the video in ways that would allow the data to be maximally useful to a wide variety of research programs addressing diverse questions. For example, existing software was modified so that researchers could easily sort video according to (a) the family members appearing on-screen, and (b) their locations inside or outside the home (e.g., kitchen, bathroom, car, restaurant). The investment of resources in software development and the employment of large numbers of research assistants who coded every second of the 1600 hours of footage has paid off thanks to the universal utility of this system and the data it generates. The effort greatly facilitated the first steps of numerous research projects by expediting the identification of relevant video segments for further coding.

Defining constructs through behavioral coding

Whether direct observation of behavior takes place in a laboratory or in the field, the constructs of interest are ultimately operationalized through the coding and analysis of behavior. One of the main advantages of direct observation is the increase in reliability that results when only one person (i.e., the researcher) is "making sense" of raw observations, rather than many participants. For example, to study work–family conflict in a direct observational study, a specific indicator of family conflict – such as hostile questioning or blaming, or pressure for change – would be explicit and applied across families without the involvement of the family members in the coding or counting process. The uniform application of a behavioral code to different families, which significantly reduces random measurement error, is assessed through measures of inter-rater reliability. There are trade-offs in behavioral coding systems between the reliability and the meaningfulness of the codes that are developed. Failure to achieve high reliability on a complex behavior that may be manifested in a variety of ways (e.g., frustration) may call for the use of a variable that can be more reliably coded (e.g., negative emotion expression).

Behavioral coding variables can represent notably different units of analysis. For example, the units analyzed by a micro-analytic video coding system like the

Social Support Interaction Coding System (SSICS; Bradbury and Pasch, 1994), which assesses turn-by-turn couple support behavior, differs from the units of analysis in a global rating system like the Couple Interaction Rating System (CIRS; Heavey, Gill and Christensen, 1998), which assesses the overall intensity of demand–withdraw behaviors in a couple interaction. With audio-recordings, a parallel issue exists regarding the scope with which transcripts of speech are coded. Some tackle this at the level of the text itself, using defined categories for grouping and quantifying words (e.g., prepositions, negative emotion; Language Inquiry and Word Count; Pennebaker, Booth and Francis, 2007). This method was applied to the CELF parents' narrations of self-guided tours of their homes in order to address questions about how home life can facilitate or impede physiological and emotional recovery after work. Wives who used words like "cluttered" or "unfinished" to describe their homes showed flatter (i.e., less healthy) diurnal cortisol slopes and increases in depressed mood over the course of an average day. The use of restful words, such as "comfortable" and "peaceful," or references to nature, were associated with steeper (i.e., healthier) cortisol slopes and decreased depressed mood over the day. Other approaches used in coding speech focus on the content or interactional process in conversations (e.g., problem-solving and monitoring in conversations about peer and academic problems). We have focused on methods that are used by psychological and quantitative researchers. Scholars outside of our field studying the same phenomena (e.g., applied linguists, conversation analysts, linguistic anthropologists) often adopt very different analytic techniques. For example, Ochs and Izquierdo (2009) examined the socialization of responsibility in childhood in the CELF data using an anthropological framework and qualitative methods that differ considerably from the methods we describe here.

The coding of naturalistic observational data differs from the coding of observations made in the laboratory, where *what* is being observed (e.g., type of interaction, actors present) and the *way* in which it is observed (e.g., camera angle, sound quality, close-ups) are tightly controlled and uniform across participants. Imagine trying to apply a system developed to code turn-by-turn interaction in couples instructed to discuss a specific topic to a stream of naturalistic observations. It is analogous to the expectation that a response scale developed for couples to respond to specific questions about their relationship (e.g., "how many times each week do you and your spouse argue?") be used to code responses to an open-ended question (e.g., "tell us a bit about your relationship"). It just isn't possible. In the real world, interactions occur alongside other activities, conversations on topics of interest to the researcher (when they do spontaneously occur) may be interrupted by another family member and fail to resume, or participants may walk in and out of a room mid-conversation, limiting the ability of the researcher to capture the full content of a conversation. Recording devices that have to move with participants may not be sensitive enough to pick up quiet or poorly articulated phrases. In addition, a particular topic, or a particular family member, may weave in and out of an interaction. As a result, rather than a continuous (and uniform) "conversation" on a single subject of interest, the relevant "talk" between two

family members may start and stop at multiple points over an extended period. Similar difficulties occur with visual recordings, such as participants' faces being in shadow, turned away from the camera, or behind an object that obscures full view. More subtle behaviors like facial expressions may be very difficult for coders to notice or agree upon, and phenomena of interest may occur at extremely low base rates, all resulting in lower reliability coefficients than are typically obtained in more controlled laboratory settings.

Naturalistic observational data call for new coding systems, often developed through an iterative process in which concepts are defined, shaped, and re-defined through repeat engagement with the raw data. A manualized coding system develops through a primarily bottom-up, data-driven process in which the researcher may be discovering – rather than imposing – a definition for a behavior or process. Coded behaviors may reflect lower order constructs that are descriptively preserved throughout analysis (e.g., ratings of the amount of touches exchanged by a couple); higher order constructs can also emerge from these lower order constructs during data analysis (e.g., touches as one indicator of expressions of emotional support).

Reconceptualizing and reinterpreting observational data

Participant reports are in many ways locked in time. Constructs, having been defined in advance of data collection, are resistant to later revision. A work–family researcher's preconceived notions about what constitutes balance, conflict, facilitation, or spillover between the job and home domains are embedded within the selection of items and the choice of wording on participant-report scales. Direct observational data, on the other hand, can be revisited by different scholars, even decades later, with new coding systems and entirely rewritten constructs. The opportunity to reconceptualize and reinterpret the raw data with fresh ideas and the promise of "future" research are probably the most overlooked benefits of direct observational research. When a study's raw data consist of ratings made by an observer – whether that observer is a family member responding to items on a questionnaire or a researcher making notes about a family behavior – it is certainly possible to return to the data and redefine concepts of interest, but the options are limited. With participants' responses to items on a questionnaire and researchers' real-time coding of live interaction, conceptual constraints are reflected in the questions that were posed to the participants or in the guidelines followed by live observers.

There is more flexibility when the raw data consist of video and audio recordings of families. For example, even though the behavior observed in a typical laboratory study is constrained and controlled to some degree, the researcher can return to the recordings with a new set of concepts or to examine different aspects of the raw data. Video, in particular, offers a storehouse of information – images, language, movement – comprising a dense encyclopedia of information that can be deciphered for multiple purposes. In the CELF study, video taken during weekday afternoons were included in the study of daily family reunions described earlier;

greetings that parents received when they arrived home were coded for constructs such as positive behaviors, distraction, and reports of information (Campos, *et al*., 2009). The same recordings were also included as part of the analysis mentioned earlier of working couples' after–work family interactions for which levels of behavioral involvement (talking, social engagement) and negative emotion expression were coded (Wang, *et al*., 2011). And a current project uses the same video data in an investigation of marital social support that includes constructs such as solicitations and offers of support, support quality, and type of support (instrumental, emotional). Direct detailed raw observations of families – such as video and audio recordings – provide an ideal medium for researchers to engage and re-engage with different questions and constructs.

Conclusion

We have argued that work–family researchers have much to gain from direct observational methods. The advantages over participant report methods include improvements in external validity, protection against correlations inflated by respondent biases (the most common source of Type I error in our field), and reduction of random measurement error (i.e., reductions in Type II error). In addition, direct observation can be used with a wider population; for example, pre-literate children can be included as participants in a family study. Perhaps most important, direct observational data can contribute to the development of key constructs and the generation of new hypotheses. One pattern of findings from the CELF investigation of links between job stress and family social behavior after work highlighted the intensity of an employed family member's reactions to others' initiations of interaction. Response intensity was coded only in response to openings for a social interaction that were presented by others, regardless of the affective quality or valence of the response; amount of talking and expression of emotion were coded as separate variables. For instance, a response to a question could range from simply ignoring it to an effusive and articulate answer (with either a positive or a negative tone) (Wang *et al*., 2011). Although our interest in social withdrawal as a response to job stress provided the impetus for examining this variable, response intensity itself is not a familiar concept in the work–family literature. It is probably too subtle and complex a variable for family members to observe and accurately describe in a participant report measure. Nonetheless, CELF's naturalistic data provided novel information about response intensity for work–family researchers by suggesting how the quality of social behavior at home is influenced by stress at work.

Unquestionably, the benefits of direct observational studies are balanced by challenges. The intensity and complexity of data collection procedures, especially when conducted in the context of a naturalistic design, translate into extra expenses and smaller sample sizes. One result is a reduction in statistical power that all but rules out the testing of small effects. That problem can be minimized to some extent by increasing the homogeneity of the sample (based on individual and family characteristics, such as SES, occupation, age of children, etc.), which

reduces variance in the data attributable to group differences. However, that solution reduces the generalizability of the study's findings to more narrow populations. Naturalistic observational studies also add analytic complexities associated with parsing and coding of the data. Fortunately, the challenges are not insurmountable. The elasticity of naturalistic observational data, particularly when collected with minimal processing by an observer (i.e., video and audio recordings) is its greatest selling point. As demonstrated by the CELF study, resources can be pooled and scholars from different disciplines and sub-disciplines can coordinate methodologies to produce a multi-layered archive of naturalistic data. (See Ochs *et al*., 2006 for a more detailed description of the video and scan sampling methods used by CELF.) Although initial investments in time and funding may seem burdensome, the picture changes drastically when the archive's many purposes and applications across different fields are considered.

Ultimately, a mixture of methods provides the most valuable and valid information about a construct and the most complete answer to a research question. Findings based on both direct and indirect observations, each flawed in its own way, will offer unique insights that contribute to a rich mosaic of knowledge in our field.

References

Bradbury, T.N. and Pasch, L. (1994) *The Social Support Interaction Coding System (SSICS)*. Unpublished coding manual, University of California, Los Angeles.

Campos, B., Graesch, A., Repetti, R.L., Ochs, E. and Bradbury, T. (2009). Naturalistic observation of opportunities for social interaction in contemporary family life. *Journal of Family Psychology*, 23(6), 798–807.

Coan, J.A. and Gottman, J.M. (2007). The specific affect (SPAFF) coding system. In J.A. Coan and J.J.B. Allen (Eds.) *Handbook of Emotion Elicitation and Assessment* (pp. 106–123). New York, NY: Oxford University Press.

Costigan, C.L., Cox, M.J. and Cauce, A.M. (2003). Work-parenting linkages among dual-earner couples at the transition to parenthood. *Journal of Family Psychology, 17,* 397–408.

Driver, J.L. and Gottman, J. M. (2004). Daily marital interactions and positive affect during marital conflict among newlywed couples. *Family Process, 43,* 301–314.

Ekman, P., Friesen, W.V. and Hager, J.C. (2002). *Facial Action Coding System* [E-book]. Salt Lake City, UT: Research Nexus.

Gottman, J. M. and Krokoff, L.J. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57*(1), 47–52.

Gottman, J.M. and Driver, J.L. (2005). Dysfunctional marital conflict and everyday marital interaction. *Journal of Divorce and Remarriage, 43*(3/4), 63–77.

Grossman, F.K., Pollack, W.S. and Golding, E. (1988). Fathers and children: Predicting the quality and quantity of fathering. *Development Psychology, 24,* 82–91.

Hayden, L.C., Schiller, M., Dickstein, S., Seifer, R., Sameroff, S., Miller, I., Keitner, G. and Rasmussen, S. (1998). Levels of family assessment I: Family, marital, and parent–child interaction. *Journal of Family Psychology, 12*(1), 7–22.

Heavey, G., Gill, D.S. and Christensen, A. (1998). *The Couple Interaction Rating System.* Unpublished document. University of California, Los Angeles.

Heyman, R.E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment, 13*(1), 5–35.

Larson, R. and Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15, 41–56.

Larson, R.W., Richards, M.H. and Perry-Jenkins, M. (1994). Divergent worlds: The daily emotional experiences of mothers and fathers in the domestic and public spheres. *Journal of Personality and Social Psychology, 67*, 1034–1046.

Mehl, M.R., Pennebaker, J.W., Crow, D.M., Dabbs. J. and Price, J.H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments and Computers, 33*(4), 517–523.

Ochs, E., Graesch, A.P., Mittmann, A., Bradbury, T. and Repetti, R. (2006). Video ethnography and ethnoarchaeological tracking. In M. Pitt-Catsouphes, E.E. Kossek and S. Sweet (Eds.), *The work and family handbook: Multi-disciplinary perspectives, methods and approaches* (pp. 387–409). Mahwah, NJ: Lawrence Erlbaum Associates.

Ochs, E. and Izquierdo, C. (2009). Responsibility in childhood: Three developmental trajectories. *Ethos, 37*(4), 391–413.

Ochs, E., Shohet, M., Campos, B. and Beck, M. (2010). Coming together for dinner: A study of working families. In K. Christensen and B. Schneider (Eds.), *Workplace Flexibility: Realigning 20th Century Jobs to 21st Century Workforce* (pg. 57–70). Ithaca, NY: Cornell University Press.

Pasch, L.A. and Bradbury, T.N. (1998). Social support, conflict, and the development of marital dysfunction. *Journal of Consulting and Clinical Psychology, 66*(2), 219–230.

Patterson, G.R. and Forgatch, M.S. (1995). Predicting future clinical adjustment from treatment outcome and process variables. *Psychological Assessment, 7*(3), 275–285.

Pennebaker, J.W., Booth, R.J. and Francis, M.E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX: LIWC (www.liwc.net).

Repetti, R.L. (1987). Individual and common components of the social environment at work and psychological well being. *Journal of Personality and Social Psychology*, *52* (4), 710–720.

Repetti, R.L. (1989). Effects of daily workload on subsequent behavior during marital interaction: The roles of social withdrawal and spouse support. *Journal of Personality and Social Psychology*, *57* (4), 651–659.

Repetti, R.L. (1994). Short-term and long-term processes linking job stressors to father-child interaction. *Social Development*, *3* (1), 1–15.

Repetti, R.L. and Wood, J. (1997). Effects of daily stress at work on mothers' interactions with preschoolers. *Journal of Family Psychology, 11,* 90–108.

Saxbe, D.E. and Repetti, R.L. (2010). No place like home: Home tours correlate with daily patterns of mood and cortisol. *Personality and Social Psychology Bulletin*, *36*, 71–81.

Saxbe, D.E., Repetti, R.L. and Graesch, A.P. (2011). Time spent in housework and leisure: Links with parent's physiological recovery from work. *Journal of Family Psychology, 25*(2), 271–281.

Schulz, M.S., Cowan, P.A., Cowan, C.P. and Brennan, R.T. (2004). Coming home upset: Gender, marital satisfaction and the daily spillover of workday experience into couple interactions. *Journal of Family Psychology*, 18, 250–263.

Shiffman, S., Stone, A.A. and Hufford, M.R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1–32.

Slatcher, R.B. and Robles, T.F. (in press). Preschoolers' everyday conflict at home and diurnal cortisol patterns. *Health Psychology*.

Stone, AA. and Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*(3), 236–243.

Story, L.B. and Repetti, R.L. (2006). Daily occupational stressors and marital behavior. *Journal of Family Psychology, 20(4),* 690–700.

Wang, S., Repetti, R.L. and Campos, B. (2011). Job stress and family social behavior: The moderating role of distress. *Journal of Occupational Health Psychology, 16(4),* 441–456.