

**INTER-OBSERVER AGREEMENT AND THE EFFECTS OF ETHNICITY ON  
OBSERVATIONAL CODING OF AFFECT**

A Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

---

In Partial Fulfillment  
of the Requirements of the Degree of  
Master of Arts

---

By

Josilyn Banks

December, 2013

INTER-OBSERVER AGREEMENT AND THE EFFECTS OF ETHNICITY ON  
OBSERVATIONAL CODING OF AFFECT

---

Josilyn C. Banks

**APPROVED:**

---

Julia Babcock, Ph.D.  
Committee Chair

---

Peter Norton, Ph.D.

---

C. "Chip" Raymond Knee, Ph.D.

---

John W. Roberts, Ph.D.  
Dean, College of Liberal Arts and Social Sciences  
Department of English

**INTER-OBSERVER AGREEMENT AND THE EFFECTS OF ETHNICITY ON  
OBSERVATIONAL CODING OF AFFECT**

An Abstract of a Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

---

In Partial Fulfillment

of the Requirements of the Degree of

Masters of Arts

---

By

Josilyn Banks

December, 2013

## TABLE OF CONTENTS

Chapter	Page
1. INTRODUCTION _____	1
Ethnic Stereotyping _____	2
Cultural Meaning in Research _____	4
The Current Study _____	5
3. METHOD _____	7
Recruitment & Participants _____	7
Procedures & Measures _____	8
Conflict Tactics Scale-2 (CTS-2) _____	9
Coders _____	9
SPAFF Coding System _____	10
4. DATA ANALYTIC STRATEGY _____	11
5. RESULTS _____	11
6. DISCUSSION _____	13
7. RESEARCH IMPLICATIONS _____	15
8. LIMITATIONS & FUTURE DIRECTIONS _____	16
9. REFERENCES _____	18

## **LIST OF TABLES**

Table 1. Demographics by Gender _____	22
Table 2. Means, SD, and F values of the Interaction between Ethnicity of Coder and Couples' Displays of Affect _____	23
Figure 1. 2-way Interaction Ethnicity of Coder by Ethnicity of Couple _____	24

## **Dedication Page**

I would like to dedicate my thesis to my parents, Jeffrey and Karen Banks and my big “sissy” Tiffany. Thank you for your unconditional love and support.

## **ABSTRACT**

Observational research has shown that ethnicity of coders can bias their coding of people of different ethnic origin. However, observational coding systems of affect assume that emotion is universal and can be reliably coded across culture. The current study examined whether the ethnicity of the coder interacts with the ethnicity of the observed participant to influence judgments of the participants' affect displays. Couples engaged in a conflict discussion and their affect displays were coded using the Specific Affect Coding System (SPAFF; Gottman, McCoy, Coan, & Collier, 1996). Although it was predicted that coders would show a positive bias towards “in-group” (ethnically matched) participants and a negative bias towards “out-group” (not ethnically matched) participants, results suggest that no such biases exist. Rather, Caucasian coders coded more neutral affect for Caucasian couples than African American coders coded for Caucasian couples. This only significant difference between coders on judgments of positive or negative codes suggests that SPAFF is relatively robust against ethnic stereotyping.

## **Introduction**

Observational research is a useful technique that provides meaningful, socially relevant information about human behavior. Several observational coding systems have been developed in recent years. The Specific Affect Coding System (SPAFF; Gottman & Krokoff, 1989) is a widely used observational coding system designed to objectively code interactions between intimate others. Although SPAFF is a reliable and widely used coding system (e.g., Carstensen, Gottman & Levenson, 1995; Chase, O’Leary, Heman, 2001; Holtzworth-Munroe et al., 2000; Jacobson et al., 1994), it requires interpretation of individuals’ behavior and a comprehension of the sequence of events, as with most observational coding systems. The SPAFF manual (Gottman & Krokoff, 1989) does not specifically address variations in culture or discuss the potential for dissecting cultural meaning in the communication of those being observed. Further, there is an assumption that coders are interpreting the behaviors of others in an objective, unbiased way and that observational research is not subject to the preconceived biases and beliefs of trained coders.

In short, SPAFF coders are assumed to be “cultural informants” that have expertise about the culture they are studying to interpret events, and an observation of a SPAFF coder “provides the meaning that only a competent observer in that culture could provide” (Gottman, McCoy, Coan, & Collier, 1996, p. 25). However, it is not clear to what extent trained coders must understand the cultures of those being observed. The question is if developing accurate cultural meaning is something that can be trained within an observational coding lab or if coders must use their inherent understanding of subcultures. It is also unclear whether coders must be matched to participants by subculture to provide accurate cultural meaning to their interpretations. If it were the case that observational

research must rely on coders' understanding of subcultures, researchers would be remiss to overlook the role of unintentional negative biases and stereotypes as potential roadblocks to gathering accurate cultural meaning during intimate interactions.

Ethnic stereotyping can unintentionally emerge during observational research; SPAFF may be subject to bias due to unintentional ethnic stereotyping on the part of the coders who observe the behavior. It is important to understand the role ethnic stereotyping plays within our society and the pervasive influence that stereotyping has in multiple areas. Thusly, we should be able to examine the potential effects of ethnic stereotyping within research labs that utilize observational research. Given that stereotypes of African American couples and Caucasian couples are most widely accessible, this project examines the how these powerful—if subtle—expectations of how African American and Caucasian couples interact influence SPAFF coding as coded by in-group and out-group (i.e., not ethnically matched) coders.

### **Ethnic Stereotyping**

Preconceptions about how people from a specific ethnic group typically behave can distort individuals' perceptions of other people's behavior; these preconceptions are often simplified and based on minimal knowledge or experience. Ethnic stereotyping adds a subjectivity that the field of observational research cannot afford. It is vital to explore ethnic stereotyping not only in a systematic way, but also in a way that can shed some light on the objectivity that is the bread and butter of observational research.

Images of ethnic stereotyping can be accessed in various media outlets and make it difficult if not impossible for the typical American to ignore. Pervasive images in society (e.g., in the television, movies, and music videos) portray African American individuals as

aggressive (Mattis & Bell, 2000). Even among youth, African American males are seen as aggressive and “tough” and African American females are seen as loud and confrontational (Morris, 2007). Comparatively, Caucasian Americans are seen as the “standard” of American culture and society. African Americans (and other minorities) are typically regarded through a comparative lens of the white majority. In this vein, African Americans are *louder, tougher, and more confrontational* than their Caucasian counterparts.

Importantly, stereotypes can be applied to individuals as well as dyads or groups. Negative stereotypes of African American couples are ubiquitous within U.S. culture. These negative expectations and views of the nature of African American relationship dyads may subtly influence coders’ interpretations of behavior (Sagar & Schofield, 1980). African American women have long-standing, stereotyped images of castrating, domineering, and insensitive figures” who must be in control (Mattis & Bell, 2000). The African American man has traditionally been equated with “danger, social chaos, and predatory violence” (Mattis & Bell, 2000). Obviously these traits would not bode well for creating mutually respectful and satisfactory marriages. Looking at these traits as traits of potential life partners sets the stage of a ticking time bomb of marital dissatisfaction and tension.

It comes as no surprise that African American male-female relationships have been categorized as “inherently affectionless and inevitably conflictual,” as well as “problematic and destructive” (Mattis & Bell, 2000; Willis, 1990). These preconceived expectations can understandably be damaging to the perceptions of African American couples within a society that expects these intimate relationships to be inherently “doomed to fail.” With regard to Caucasian couples, stereotypes are often positively valenced. Caucasians are often portrayed in the media as exhibiting more positive affect in their communication styles with more

frequent use of apologies, agreements, and empathetic understanding. The question of whether or not SPAFF is robust against such pervasive negative expectations of the nature of African American relationships as well as the positive expectations of the nature of Caucasian relationships.

### **Cultural Meaning in Research**

One would expect that in-group raters might view members of their in-group as more normative and similarly observe less negative exchanges. These perspectives rest on the assumption of different belief and meaning systems across cultures (Albright et al., 1997). If this assumption is untrue, then there should be one universal pattern of associations between observations and external cues in all affective codes and thus observational coding should result in objective ratings. If these premises do not hold true, then these kinds of pervasive stereotyped perceptions may especially affect out-group coders' ability to rate observed behavior objectively on the SPAFF coding system. This lack of objectivity has the potential to undermine the reliability and utility of observational research.

Research by Albright et al. (1997) suggests that people perceive members of an out-group as relatively homogeneous. On the basis of this research, when making judgments of individual members of another culture, people should simply use their stereotypes about the out-group, a process that would yield a lack of differentiation of the targets (i.e., all targets would be judged by the stereotypes of that group). Given the pervasiveness of negative stereotypes of African Americans when compared to Caucasian Americans, one possible outcome of observational coding influenced by cultural bias is that in-group and out-group associations may have a weaker than expected influence, particularly with African American observations of Caucasian interactions. Given pervasive images of Caucasian individuals and

their interactions as the standard of family interactions (Kochman, 1981), African American coders may view Caucasian interactions as inherently more normative and positive when compared to African American couple interactions. In this vein, African American coders may be habituated to viewing Caucasian interactions in a more positive manner than they view interactions between African American couples if Caucasian interactions are viewed as the gold standard of healthy relationships.

Another possibility is in different interpretations of couples' interaction when there is open disclosure. According to Aschenbrenner (1975), African Americans have a greater tolerance for open, intense disclosure than Caucasian couples (Kochman, 1981). If the Caucasian couples engaged in more mutual expression and exchange of ideas and opinions in the lab paired with a higher intensity of discourse (i.e., louder voice of more expressive body language), African American coders may rate these exchanges as more positive or normal, whereas the Caucasian coders may view them as neutral or negative.

### **The Current Study**

Observation of family interaction has become a widespread procedure used within the field of psychology. However, to date, there has been little systematic investigation of interpersonal perception and how it may be biased by cultural stereotypes and whether or not the rater belongs to the ethnic group that he/she is asked to judge. Bias is defined here as systematic variation in ratings based on attributes of raters rather than on the attributes of the persons being rated (Melby, Hoyt & Bryant, 2003). I focus specifically on race-related bias: that is, bias attributable to the race of the rater, or to raters' differential perceptions of same- and different-race targets.

All observational coding systems should consider inter-observer agreement and the potential for bias among independent raters stemming from differences in ethnic background. Discrepancies between coders could have important implications for researchers who seek to code the behavior of diverse ethnic groups. Ethnic stereotyping may interfere with the coders' ability to be objective while rating the behavior of an individual from a different ethnic group, although there has been no research to support this theory within the observational research literature. For example, Gonzales and colleagues (Gonzales, Cauce, & Mason, 1996) found that non-African American coders rated African American mothers as significantly more controlling and rated the mother-daughter dyads as significantly more conflictual than did African American coders. They recommended that coders be matched to participants by ethnicity to avoid these subtle ethnic biases. To date, no study has examined whether SPAFF, in particular, is similarly subject to ethnic bias and if cultural informants must be ethnically matched to the research participants. This study seeks to explore the interaction of ethnicity of both coder and participant rather than simply the main effect of the ethnicity of the coder.

The purpose of this study was to observe the extent to which ethnic biases may influence coding of observed affect. While emotional displays may be universal (Ekman & Friesen, 1971; Scherer, Wallbott, & Summerfield, 1986), their meaning and impact may be judged differently by different coders within the context of social interaction.

### **Hypotheses**

1: In-group observers will code more positive affect (such as humor, interest, and validation) among ethnically matched participants than out-group coders.

- 2: In-group observers will code less negative affect (such as stonewalling, defensiveness, disgust) among ethnically matched participants than out-group coders.
- 3: In-group observers will code less aggressive behavior (such as domineering, belligerence, and contempt) than out-group coders.
- 4: In-group observers will code less distressed affect (such as fear, tension and sadness) among ethnically matched participants than out-group coders.
- 5: There will be no significant difference in the coded neutral affect between in-group and out-group coders. In short, in-group observers are expected to code their own ethnicity in a more favorable light and in a less aggressive manner.

## **Method**

### **Recruitment**

Participants were couples recruited from the ethnically diverse Houston area. Participants were recruited as part of a larger study on emotional regulation in domestically violent and nonviolent couples. All participants were recruited from the community via local newspaper ads and flyers requesting couples who were married or living together as if married for at least 6 months, 18 years of age, and able to speak and write English proficiently. Female participants were contacted by phone by trained undergraduate interviewers who administered the violence subscale of the Conflicts Tactics Scale (CTS; Straus 1979) to determine eligibility in the study.

### **Participants**

In the larger study of 110 couples, 40% were African American, 19% Hispanic and 33% Caucasian and 8% were from other racial or ethnic origins. Approximately 22% of the couples were interracial. For this project, Hispanic, interracial, and “other” ethnicity couples

were excluded and only African American (N = 29) and Caucasian couples (N= 24) were included. Table 1 shows the demographic break-down by ethnicity and gender. Average age of the sample was 31 (SD = 9.22). Median gross family income was approximately \$39,000 per year (SD = 40,997). Average length of the relationship was 5 years (SD = 4.34). There were no significant differences on age, income, and length of relationship. The larger study was a study primarily focused on violent couples, the couples from this sample reported, on average of 4.3 reported incidents of violence (SD = 1.72). Of note, female reported physical violence of African-American and Caucasian participants was significantly different  $F(1, 53) = 4.70, p = .04$ , with African American couples reporting more lifetime physical assault.

### **Procedures**

During one assessment period that lasted approximately 3.5 hours, couples were separated to complete a series of questionnaires and then reunited to participate in a conflict discussion. Research assistants administered the Play-by-Play Interview (Hooven, Rushe & Gottman, 1996) in order to clarify an actual conflict area in their relationship. Couples were then asked to engage in a 15-minute conflict discussion while their displayed affect was coded. Couples were paid \$80 to \$100 for their participation; couples received \$10 extra if they came on time for their first scheduled appointment.

### **Measures**

Participants were asked to identify the most significant areas of disagreement in their current relationship. Research assistants administered the Play-by-Play Interview (Hooven, Rushe, & Gottman, 1996) in order to clarify an actual conflict area in their relationship. The couple was then asked to discuss the disagreement they rated as most important in a 15-minute videotaped discussion while their displayed affect was coded. Per the SPAFF manual,

two video cameras were used to obtain clear images of each participant's face and top of the torso to optimize the ability to observe facial expression and body language. The two images were then recorded in a split-screen format so that partners appeared side-by-side facing the screen. To minimize the influence of one partner's behavior on the coding of the other partner, one half of the split video screen was covered so that only one partner was visible at a time, but no effort was made to block out the other partner's vocalizations.

**Conflict Tactics Scale-2 (CTS-2).** The CTS-2 (Straus, Hamby, Boney-McCoy, & Sugarman, 1995) is considered the gold standard to assess domestically violent behavior within the past year. The CTS-2 is a 78-item self-report questionnaire that assesses the severity and frequency of physical, sexual, and psychological abuse committed by intimate partners. Five scales measure negotiation, psychological aggression, physical assault, sexual coercion, and injury. The CTS-2 was used to determine final group membership into nonviolent and domestically violent groups based on both male and female reports as well as determine frequency of men's IPV based on male and female reports. Internal consistencies on the CTS-2 ranged from .49 to .78.

**Coders.** Six coders coded the 53 videos in question. Each 15-minute conflict discussion was coded by one Caucasian and one African-American student trained in using the Specific Affect Coding System (SPAFF; Gottman, McCoy, Coan, Collier, 1996). All coders were female, ranging in age from 24 to 48 years old. Three were undergraduate students, and three were post-baccalaureate students. All were trained to reliability on SPAFF using training tapes. Inter-rater reliability on the training tapes ranged from Kappa = .74 to .96.

**SPAFF Coding System.** SPAFF categorizes 16 emotions based on facial affect, vocal tone, body language, and content of speech. The Specific Affect Coding System (SPAFF) is the most fully developed observational coding system designed to code universal emotions of conversations between family members to date (SPAFF; Gottman, McCoy, Coan, & Collier, 1996). SPAFF synthesizes changes in vocal tone, body movement, facial expressions and content into labeled specific emotions. The 169-page SPAFF manual sets rules for coding both verbal and nonverbal information into 16 discrete variables including emotions such as anger and joy. SPAFF coders are trained to code videotaped marital interactions continuously, while judging which of the 16 mutually exclusive variables (including “neutral”) is present for each instance of tape (Waldinger, Schulz, Hauser, Allen, Crowell, 2004). The manual outlines separate categories for 10 negative and 5 positive emotions. These 15 codes were collapsed into four categories plus neutral in the current study.

The following categories were created to account for the lack of sufficient variability within some of the less frequent codes in the current sample: **Positive** affect (interest, humor, affection/caring, joy/surprise, validation); **Negative** affect (stonewalling, defensiveness, and disgust); **Aggressive** affect (anger, domineering, contempt, and belligerence); **Distressed** affect (fear/tension, whining, and sadness); and **Neutral** affect (neutral, non-codable behaviors). Videotapes of the conflict discussion were coded on a second-by-second basis using the Video Coding Station (Long, 1998). In order to calculate reliability, a second coder coded 25% of the interactions. Inter-rater reliability for the specific codes ranged from Kappa = 0.67 to 0.99 ( $p < 0.001$ ). Kappas for the codes of positive, negative, aggressive, distressed and neutral, were .87 .89, .67, .72, and .74 respectively.

## **Data Analytic Strategy**

All hypotheses were tested using a 2 x 2 x 2 (ethnicity of coder x ethnicity of couple x gender of partner) repeated measures MANOVA conducted to determine whether there were statistically significant differences in SPAFF codes when looking at the interactions between the ethnicity of the coder and the ethnicity of the couple. For this analysis, ethnicity of couple and ethnicity of coder were the independent variables, while the SPAFF ratings—as coded by Caucasian and African American coders—acted as the dependent variable. The ethnicity of the coder and 5 SPAFF codes acted as the repeated measures. Furthermore, following the omnibus *F* test, univariate analyses were examined; Type I error was addressed through LSD corrections to the alpha level of significance.

## **Results**

No hypotheses were made for the main effects due to ethnicity of coders or ethnicity of the participant. Rather, this study explored the interaction between the ethnicity of coder and participants. SPAFF codes were not found to be normally distributed across ethnicity of coder and couples. The assumption of sphericity was violated, as assessed by Mauchly's Test of Sphericity,  $X^2(9) = 92.875, p < .01$  &  $X^2(9) = 67.828, p < .01$ , therefore degrees of freedom were corrected for using Greenhouse-Geisser estimate of sphericity ( $\epsilon = .520$  &  $.610$ ). While no hypotheses about the interaction between ethnicity of coder, ethnicity of couple, and gender of subject were made, these 3-way interactions were explored.

The results show that there was a significant 2-way interaction between ethnicity of code and ethnicity of couple on the SPAFF coding: omnibus  $F(8,406) = 2.56, p = .01$ .

Examination of the univariate tests of significance indicate an interaction between ethnicity of coder, and ethnicity of the couple for only one affect code: neutral. Examination of the means in Table 2 indicates that Caucasian coders rated more neutral affect among Caucasian couples ( $M = 69.33, SD = 9.76$ ) than did African American coders among Caucasian couples ( $M = 25.00, SD = 8.53$ ). Figure 1 shows the 2-way interaction of ethnicity of coder by ethnicity of couple for all affective codes. None of the 3-way interactions were significant.

Contrary to the original hypothesis, Caucasian coders and African American coders did not differ in their observations of in-group and out-group couples on positive, negative, aggressive, or distressed affect. There were no significant interactions for any of the four other affective codes of interest. Contrary to the first hypothesis, in-group coders did not code more positive affect than out-group coders:  $F(2,50) = .322, p = .726$ . Both Caucasian and African American coders observed in-group and out-group couples to have similar representations of positive affect. Of note, there were also no significant differences between in-group and out-group coders observations of negative affect  $F(2,50) = 1.969, p = .150$ . Coders were not more likely to code in-group couples as less negative than out-group coders. Likewise, coders did not observe more negative affect during interactions between out-group couples. There were also no significant differences between in-group and out-group coders observations of aggression  $F(2,50) = .773, p = .467$ . Coders did not observe couples to be particularly more aggressive when they were members of the coder's ethnic out-group. Finally, contrary to the fourth hypothesis, there were no differences between in-group and out-group coders observations of distress  $F(2,50) = .431, p = .652$ . Coders were not more likely to view out-group couples as more distressed than in-group couples. The effect of couples' ethnicity accounted for the majority of variance.

## Discussion

Given this field's interest in understanding interactions between intimate partners, it is important to develop and use effective and culturally valid rating systems. SPAFF has been proven to be ecologically valid, however, there has been no study of its cultural validity. As cultural diversity increases, so does the role that culture takes in research, particularly in observational research. Because expressions and perceptions of behaviors (e.g., communication, warmth) may differ across cultural or ethnic groups, it is important to examine variations in research assessments within and across such groups (Archer, 1997; Bailey, Nowicki, & Cole, 1998; Hines, 1993; Lewis-Fernandez & Kleinman, 1994; Okazaki & Sue, 1995). In this study, ethnicity of coders and the ethnicity of the observed couples interact to impact observations of emotion when intimate partners engage in a conflict discussion.

The significant 2-way interaction of ethnicity of coder by ethnicity of couple supports the notion that in general, SPAFF is robust against unintentional ethnic stereotyping. The findings of the present study suggest that contrary to the original hypotheses, our coders did not code in-group couples more favorably than out-group couples. Unlike previous studies, (Albright et al., 1997; Phelps, Meara, Davis & Patton, 1991), no biases were observed with regard to in-group and out-group coding of negative affect in general. This is promising news for those training observational researchers employing SPAFF with diverse samples. Even though the SPAFF manual does not teach cultural sensitivity, it appears that the lack of attention paid to cultural sensitivity training in the context of SPAFF training does not appear to impact the variability of scores across in-group and out-group coders.

However, it does appear that in-group Caucasian coders are better at recognizing or simply more likely to code neutral affect. Caucasian coders more frequently observed neutral affect among Caucasian couples than African American coders did. Although findings of this study only hold true for observations that Caucasian coders made of in-group couples, outside observers may bring greater objectivity and standardization to assessments of interaction processes (Melby, Hoyt & Bryant, 2003). Out-group coders may “over code” behaviors that in-group coders identify as neutral, attributing meaning to another emotion. Likewise, in-group coders may “under code” behaviors that out-group coders believe can be explained by other codes as these behaviors may be observed more frequently in in-group interactions. However, African American coders also showed the same pattern of neutral coding among in-group couples, although they tended to code fewer instances of neutral affect overall. Of note, the data did not show any evidence of a systematic way of attributing neutral affect. This pattern of ratings may be due to culturally mediated differences in interpreting behavioral cues (e.g., Albright et al., 1997; Matsumoto, 1989; Phelps et al., 1991). When observing Caucasian couples, African American coders may not have interpreted affect that Caucasian coders found to be neutral in a systematic way, however, these coders observed this same affect to be meaningful in a different way than Caucasian coders did.

Regarding the methods of the current study, the topic of conversation may have influenced the type of codes observed by in-group and out-group coders alike. Given the topic of the discussion was one of agreed upon conflict/disagreement, the original supposition that Caucasian coders may code more neutral affect in situations where there is more mutual exchange of ideas coupled with more intensity in tone, volume and physical

expressiveness may shed some light on these findings. It is possible that given a lower tolerance of open, intense discussion, Caucasian coders may have been more likely to code neutral affect during these moments than African American coders. Nonetheless, these results only hint at some larger explanation of group differences.

## **Research Implications**

These data suggest that SPAFF (Gottman, McCoy, Coan, & Collier, 1996) is a reliable observational coding system and can be generalized cross-culturally. The reliability of this interpersonal observational coding system was determined based on Caucasian coders rating Caucasian couples. Importantly, this research shows that SPAFF is robust when in-group and out-group coders are coding both Caucasian and African American couples. Given SPAFF's wide usage in observational coding research, this speaks to SPAFF's generalizability amongst ethnically diverse samples.

This research focused on a set of questions concerning consensus (or agreement) in ratings, which is a necessary but not sufficient condition for accuracy (Kenny, 1994). Future research could benefit the field by focusing on questions that center on accuracy (or validity). Accuracy of partner affect can be assessed in many ways, including comparison of coder ratings with criterion ratings, comparison of coder ratings with family-member and/or self-reports, and examination of predictive validities (Melby, Hoyt & Bryant, 2003). Family-member and self-reports may provide further insight into the meaning behind subtle affective cues among couple interactions. This study assesses the generalizability of codes when ethnicity and gender are accounted for. Assessing generalizability offers an attractive approach to the study of ethnicity of coder and ethnicity of participant effects in

observational ratings. There is also the potential for increasing understanding of culture-related dynamics of the couples observed.

### **Limitations & Future Directions**

Although the ethnically diverse sample of both participants and trained coders is a rarity in marital interaction research, the current study utilizes only one sample of participants and a small sample of coders. The labor- and time-intensive nature of SPAFF training and the desire to insure that all coders were trained under the same protocol also limited the number of coders. The small number of raters in the present study limits the generalizability of the results. Smith (1978) suggests that reliable estimates of coder ethnicity variance may require eight raters per group, rather than the three (European-American) or three (African-American) studied here. It is not clear to what extent these findings can be generalized to other samples in other research labs that use the SPAFF coding system. There also is a need for replication using dyads representing additional racial or ethnic groups. For example, how similarly might the Caucasian and African-American observers rate couples from a third group (e.g., Asian American, Hispanic)? Future researchers may want to investigate similarities in out-group and in-group coding of different ethnicities.

Previous research has found that in-group/out-group consensus increases with increased training of the coders (Melby, Hoyt & Bryant, 2003). Coders in this study ranged in experience, but all coders had at least 3 months of experience coding with SPAFF. Future studies could examine the effects of SPAFF experience on inter-observer agreement. Additionally, examining how naïve coders compare to expert coders would provide more insight into the role that perceptions of cultural meanings may have on SPAFF codes. Over time, naïve coders may come to rely less on their pre-conceived beliefs about out-groups.

Assessing behaviors in a single type of interaction task (i.e., a conflict task) may have limited the range of behaviors elicited. Other laboratory interaction tasks, such as discussing event of the day or positive events, are likely to generate more positive exchanges and ethnic differences in how affect is coded may emerge during these tasks. Future research could benefit from using additional types of interaction tasks, as well as from comparing observer ratings with family member self-reports (Gonzales, Cauce, & Mason, 1996; Hayden et al., 1998). Of note, the sample studied is a sample selected based on self and partner report of physical violence. Differences may be observed when looking at interactions between non-violent couples.

The current study seeks to add to the limited amount of research in the area of ethnic diversity, observational coding, and couple's interactions, indicating that the coding of displayed affect using SPAFF is relatively robust to ethnic bias. Researchers are encouraged to apply SPAFF to diverse populations provided that their coders are well-trained. Although SPAFF training does not appear to require detailed discussion of culture as part of the training module, coders may benefit from exposure to culturally diverse training videos. In this way, issues with inter-rater reliability can be addressed in the training phase rather than during data collection. Providing an accurate cultural meaning from interactions within and across cultures continues to be a goal of observational research of intimate partners. As we continue to seek a better understanding of the role that culture, and ethnicity play in observation and interpretation of partner interactions, researchers must continue to ask questions that are important, even if they are uncomfortable.

## References

- Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., Winquist, L., & Yu, D. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology*, 72, 558–569.
- Archer, D. (1997). Unspoken diversity: Cultural differences in gestures. *Qualitative Sociology*, 20, 79–105.
- Cartensen, L.L., Gottman, J.M., & Levenson, R.W. (1995). Emotional behavior in long-term marriage. *Psychology and Aging*, 10, 140–149.
- Bailey, W., Nowicki, S., Jr., & Cole, S. P. (1998). The ability to decode nonverbal information in African American, African, and Afro-Caribbean, and European American adults. *Journal of Black Psychology*, 24, 418–431.
- Chase, K.A., O’Leary, K.D., & Heyman, R.E. (2001). Categorizing partner violent men within the reactive-proactive typology model. *Journal of Consulting and Clinical Psychology*, 69, 567–572.
- Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.
- Elder, G.H. (1973). Racial conflict and learning. *Sociometry*, 34, 151–173.
- Gonzales, N.A., Cauce, A.M., Mason, C.A. (1996). Interobserver agreement in the assessment of parental behavior and parent-adolescent conflict: African American mothers, daughters, and independent observers. *Child Development*, 67, 1483–1498.
- Gottman, J.M., McCoy, K., Coan, J., & Collier, H. (1996). The Specific Affect Coding System (SPAFF). In J.M Gottman (Ed.) *What predicts divorce? The measures*. Mahwah, NJ: Lawrence Erlbaum.

- Gottman, J., McCoy, K., Coan, J., & Collier, H. (1996). The specific affect coding system (SPAFF). In J. Gottman (Ed.), *What predicts divorce? The measures* (pp. 1–169). Hillsdale, NJ: Erlbaum.
- Hayden, L. C., Schiller, M., Sameroff, A. J., Miller, I., Keitner, G., & Rasmussen, S. (1998). Levels of family assessment: Family, marital, and parent–child interactions. *Journal of Family Psychology*, 12, 7–22.
- Hines, A. M. (1993). Linking qualitative and quantitative methods in cross-cultural survey research: Techniques from cognitive science. *American Journal of Community Psychology*, 21, 729–746.
- Holtzworth-Munroe, A., Meehan, J.C., Herron, K., Rehman, U., & Stuart G.L. (2000). Testing the Holtzworth-Munroe and Stuart batterer typology. *Journal of Consulting and Clinical psychology*, 68, 1000-1019.
- Hooven, C., Rushe, R., & Gottman, J.M. (1996). The Play-by-Play Interview. In J.M. Gottman (Ed.) *What predicts divorce? The measures*. Mahwah, NJ: Lawrence Erlbaum.
- Islam, M.R., & Hewstone, M. (1993). Intergroup attributions and affective consequences in majority and minority groups. *Journal of Personality and Social Psychology*, 64, 936-950.
- Jacobson, N.S., Gottman, J.M., Waltz, J., Rushe, R., Babcock, J.C., & Holtzworth-Munroe, A. (1994). Affect, verbal content, and psychophysiology in the arguments of couples with a violent husband. *Journal of Consulting and Clinical Psychology*, 62(5), 982-988.

- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kochman, T. (1981). *Black and white styles in conflict*. Chicago: The University of Chicago Press.
- Lewis-Fernandez, R., & Kleinman, A. (1994). Culture, personality, and psychopathology. *Journal of Abnormal Psychology*, 103, 67–71.
- Mattis, C.C., & Bell, J. (2000). The importance of cultural competence in ministering to African American victims of domestic violence. *Violence Against Women*, 6, 515–533.
- Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross Cultural Psychology*, 20, 92–105.
- Melby, J. N., Hoyt, W. T., & Bryant, C. M. (2003). A generalizability approach to assessing the effects of ethnicity and training on observer ratings of family interactions. *Journal Of Social And Personal Relationships*, 20(2), 171–189.
- Morris, E. (2007). “Ladies” or “Loudies”? perceptions and experiences of black girls in classrooms. *Youth & Society*, 38(4), 490–515.
- Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, 7, 367–375.
- Pettigrew, T.F. (2001). Personality and sociocultural factors in intergroup attitudes: A cross-national comparison. In M.A. Hogg, D. Abrams (Eds), *Intergroup relations: Essential readings. Key readings in social psychology*. (pp.18–29). Philadelphia, PA: Psychology Press/Taylor and Francis.

- Phelps, R. E., Meara, N. M., Davis, K. L., & Patton, M. J. (1991). Blacks' and whites' perceptions of verbal aggression. *Journal of Counseling and Development*, 69, 345–350.
- Saga, H.A. & Schofield, J.W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39, 590-598.
- Scherer, K.R., Wallbott, H.G. & Summerfield, A.B. (1986). *Experiencing emotion: A crosscultural study*. Cambridge: Cambridge University Press.
- Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family*, 41, 75-88.
- Tajfel, H. (1982). A social psychology of intergroup relations. *Annual Review of Psychology*, 17, 833-840.
- Waldinger, R. J., Schulz M. S., Hauser S. T., Allen J. P., Crowell, J. A. (2004). Reading others' emotions: The role of intuitive judgments in predicting marital satisfaction, Quality, and Stability. *Journal of Family Psychology*, 18, 58-71.
- Willis, J.T. (1990). Some destructive elements in African-American male-female relationships. *Family Therapy*, 17, 139-147.