THE IMPACT OF DEMOGRAPHIC CUES IN VIDEO-BASED SITUATIONAL JUDGMENT ITEMS

By

Juliya Golubovich

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Psychology

2011

ABSTRACT

THE IMPACT OF DEMOGRAPHIC CUES IN VIDEO-BASED SITUATIONAL JUDGMENT ITEMS

By

Juliya Golubovich

Situational judgment tests (SJTs) present test takers with hypothetical work-related scenarios and offer options of possible responses. Such tests are often video-based; test takers watch videos in which actors act out each scenario. This study examined the implications of actor casting choices. A relational demography perspective, which considers how demographic similarity between individuals influences the quality of their interactions with one another and their individual outcomes, was applied to examine the effects of actor demographics in a video-based SJT on test takers' responses and reactions. The race and gender of the actor participants were asked to imagine interacting with were varied to film six different versions of each of four SJT scenarios. Participants viewed one of the six versions of the test and indicated their likelihood of responding to each scenario in a number of different ways. Participants also answered questions about their reactions to the test. Findings showed that test takers did seem to attend to the race and gender of those they "interacted with" in SJT scenarios and this influenced their responses to and actual performance on the test. Actor demographics had little influence on test takers' reactions to the test, however. Implications of the findings and limitations are discussed.

ACKNOWLEDGMENTS

There is a whole army of wonderful people who helped to make this project possible and it gives me great pleasure to have this opportunity to thank them. First of all, I would like to thank my advisor, Dr. Ann Marie Ryan, for her steadfast guidance, timely feedback, and encouragement throughout this process. I would also like to thank my other committee members, Dr. Neal Schmitt and Dr. John Schaubroeck for their critical feedback and helpful insights. I would like to express my gratitude to Proctor & Gamble, and particularly Dr. Robert E. Gibby, for offering the financial support that made this study possible.

I am also extremely grateful to the 28 individuals who acted in my videos for responding to my call for help and doing an excellent job, and to the faculty members, students, and friends who helped me seek them out. I also want to thank the forty individuals who served as judges of my situational judgment items and actively helped me recruit additional people for the task.

I would like to acknowledge the team at WKAR, Michigan State University's on-campus TV station, for their professional directing, filming, and production work and Gary Schrock for his technical assistance in getting all the videos reformatted and set up for the experiment. A special thanks to my exceptional team of research assistants, Jessie Olien, Abraham Kim, and Julie Klopp, who I could always depend on two hundred percent during the data collection process, and to Leslie Baldwin for her speedy and always friendly assistance with all issues related to the participant pool.

I want to offer heartfelt thanks to my wonderful mom and older brother for their continuous moral support and earnest attempts to understand what I was trying to do. Thank you, mom, for the 24 pound care packages which had the UPS guy huffing and puffing on the way to my door.

Finally, thank you Alex, for listening patiently to my kvetching, ordering me takeout when my food supplies dwindled, and for being there for me always.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Situational judgment tests (SJTs), which present individuals with hypothetical work-related situations, are a popular tool in the realm of personnel selection and have a fairly long history of use going back to the 19th century (Whetzel & McDaniel, 2009). As such, many issues pertaining to their use (e.g., validity, response format, methods of scoring, subgroup differences) have been subjected to considerable research attention. With the advent of video-based versions (as compared to the typical paper-and-pencil format) of such instruments, however, new questions arise and have yet to be explored. One such question, and one which to my knowledge has not heretofore been examined, pertains to the impact that demographic cues in video-based SJTs might have on how test takers construe the situations presented.

Social categorization is known to influence how individuals perceive others and act in given situations (e.g., Tajfel, Billig, Bundy, & Flament, 1971; Carli, 1989). When little information about others is available, individuals use social category information to form impressions about these target individuals (Kunda & Thagard, 1996). Subsequent behavior toward these targets (and more generally) is influenced by initial impressions. It is therefore important to consider the influence of social categorization in selection settings. That is, how does social categorization influence applicants' perceptions and behaviors? Social categorization may come into play when applicants view a recruitment video featuring "employees" of particular demographic backgrounds, perceive the demographics of an interviewer, take note of the demographic characteristics of fellow applicants going through an assessment center, and read test items with male or female names. These are just several examples of recruiting or testing situations in which social categorization has the potential to influence applicant perceptions and behaviors.

Video-based situational judgment items provide yet another context that is susceptible to influences of social categorization. Test takers see the demographics (race, gender, age) of the actors in the video and use this social category information to form impressions about them. Given that responses to SJT items depend on how the items are interpreted (Brooks & Highhouse, 2006), demographic cues of stimulus actors may affect test takers' performance. To the extent that respondents' own race and gender help to determine the influence of actors' demographics, as relational demography research would suggest they would, subgroup differences in item responses may result. The aim of the study is to add to the research on the influence of social categorization in selection settings by exploring the question of how the race and gender of the actors in video-based SJT scenarios relative to the race and gender of the test takers influence responses to these scenarios. Aside from performance, applicant reactions are explored as well. These issues are investigated using two common types of SJT scenarios.

The study complements and informs existing research pertaining to the SJT and test bias, and potentially offers practitioners some much-needed guidance in choosing actors for their video-based tests. As it stands, test developers are guided in their casting choices by the desire to project a positive (i.e., diverse) image of the hiring organization but have little to no research data on what implications these choices have for the performance or reactions of test takers from different groups.

I begin the paper with a brief review of existing research on paper-and-pencil based SJTs and their more recent variant, the video-based SJT. I then introduce the relational demography perspective, which was used as the springboard for making predictions about the effects of actors' demographic cues on responses of test takers from different groups, and go on to briefly review the research on race and sex effects in the types of situations of interest in this study.

Findings from this situation-specific research will in some cases depart from the relational demography perspective that I will be using as the point of departure. I will point this out where applicable. I wrap up the theory-based portion of the paper with a brief review of the relevant research on applicant reactions and thereafter delve into describing my method of carrying out the study. The results of the study are then presented and the implications and limitations of the findings are discussed. Future directions for this line of research are suggested.

Situational Judgment Tests

Situational judgment tests (SJTs) are "measurement methods that may be used to assess a variety of constructs" (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). SJT items proffer hypothetical scenarios of problems that may arise in a work setting and options of possible responses to these situations. The situations presented may reflect actual situations that could occur on the job that the test was designed to reflect, or situations that are designed to be psychologically analogous to actual job situations (Chan & Schmitt, 1997). The premise underlying asking individuals to respond to hypothetical situations is that intentions predict future behavior (Latham, Saari, Pursell, & Campion, 1980).

SJT response instructions. After each situation is presented, the examinee is asked to indicate how he or she would respond if confronted with the given situation. Response instructions to SJTs can have either a knowledge or a behavioral tendency format (Lievens, Peeters, & Schollaert, 2008). With a knowledge format, examinees are asked to provide the best answer; with a behavioral tendency format, they are asked to choose their own most likely response. SJTs with a knowledge format are akin to cognitive ability or job knowledge tests in that examinees are asked to show that they know what the most effective response to a particular situation would be (Lievens et al., 2008). On the other hand, SJTs with a behavioral tendency

format are similar to personality tests in that examinees are instructed to demonstrate what their own behavior would be in a given situation (Lievens et al., 2008).

SJT development process. SJT development usually involves three stages. Lievens, Peeters, and Schollaert (2008) detail the process, drawing partly on the steps described by Motowidlo, Dunnette, and Carter (1990) in their work. In the first step, a job analysis is conducted, during which subject matter experts (e.g., supervisors) provide the SJT developers with instances of critical incidents (i.e., instances of behavior that reflect effective and ineffective performance) at work. Then, test developers proceed to group these critical incidents into particular elements of performance, and to write descriptions of task situations based on the incidents. In writing these situations, developers try to keep to a certain length and format. In the second step of SJT development, a new group of subject matter experts or novice employees are instructed to come up with possible responses for each job scenario written by the developers. While subject matter experts are able to generate the optimal responses for the various scenarios as well as some inferior ones, inexperienced employees can contribute alternatives along a relatively wider continuum of effectiveness. Armed with a range of response alternatives for the job situations, SJT developers can select for inclusion on the test response alternatives that fall along a continuum of effectiveness. The third and final step of SJT development is the creation of a scoring key.

SJT scoring. Scoring keys for SJTs are typically developed via one of the following three methods: empirical, theoretical, or expert-based (Bergman et al., 2006). When SJT items are empirically scored, a criterion measure (e.g., job performance) is chosen and items or response options are scored according to how well these items or response options are able to distinguish high from low scorers on the criterion measure (Lievens et al., 2008). This requires

giving SJT items to a sample of individuals to complete. Response options that are endorsed most frequently by individuals exhibiting a high score on the criterion measure are scored as correct, while options endorsed most frequently by low scorers on the criterion measure are scored as incorrect (Lievens et al., 2008). One of the obvious challenges inherent to this method of key development is that empirical keys are dependent upon the quality of the selected criterion (Campbell, 1990a, b as cited by Bergman et al., 2006).

In the case of theoretical scoring of SJTs, theory serves as the basis for the writing of items and options, or for the scoring of chosen options on a test (Bergman et al., 2006). Options that are consistent with theory are considered correct; options that are inconsistent with theory are considered incorrect. Mumford and colleagues (2008) used this method to create their Team Role Test. The test was designed to assess individuals' knowledge of team roles and of the situations when these roles are appropriate to use. The researchers wrote scenarios relevant for each team role (e.g., contributor, critic) defined by Mumford and colleagues (2006). Each scenario was designed to be most amenable to assuming one role over the others. For a given scenario, behaviors consistent with the role considered by the researchers to be most appropriate were considered correct answers, and behaviors inconsistent with the targeted role were considered incorrect answers. Olson-Buchanan and her colleagues (1998) created a key to score their conflict resolution SJT using Keenan and Olson's (1991, as cited in Olson-Buchanan et al., 1998) model prescribing what problem-resolving actions managers should take based on three dimensions of the conflict situation (i.e., short-term, long-term, interpersonal). Analogous to the empirical scoring method's reliance on criterion quality, the theoretical scoring approach must rely on the quality of the theory used.

Expert-based scoring involves asking either subject matter experts to select the best and worst response option for each item (e.g., Banki & Latham, 2010), or asking novices and experts to select just the optimal response for each item. In the latter case, options that experts select most often are considered correct, irrespective of how often novices choose these same options. In the meantime, options that novices, but not experts, choose most often are considered incorrect (Bergman et al., 2006).

An alternative to using just one of the three methods of developing scoring keys reviewed above is to use a hybrid of two keys that have been developed separately (see Olson-Buchanan et al., 1998 for an application of a hybrid method). This method is discussed by Bergman and her colleagues (2006). In one of a number of possible instantiations of the hybrid approach, one key serves as the primary key while the second key is used to remove zeros (assigned to responses that are neither correct nor incorrect for a given item) that result from the application of the primary key. Combinations of empirical keys with theoretical keys have the potential to address concerns about the individual methods (i.e., insufficient or excessive reliance on either theory or empiricism) but flaws that might be present in the separate keys are unfortunately not expunged when keys are combined.

Mode of SJT administration. An alternative to paper and pencil SJTs are SJTs that are video-based. In the latter case, hypothetical scenarios are presented to individuals visually, via video. Actors act out the situations in the actual setting of the job (e.g., Weekley & Jones, 1997) or a setting that resembles the actual setting (e.g., Dalessio, 1994). The video can be viewed by a group of test takers simultaneously or individually on computer screens. Test takers choose a response to each situation from a set of response options during the pause that is built into the video following each vignette. Video-based SJTs typically incorporate into the video a narrator

who may provide background information, instruct test takers on how to respond to the items, and perhaps even introduce each scenario (e.g., Dalessio, 1994; Jones & DeCotiis, 1986; Weekley & Jones, 1997).

Video-based SJTs have a number of advantages over their written counterparts. Because this format allows for the presentation of both verbal and visual cues (Jones & DeCotiis, 1986), video-presented scenarios appear more detailed and realistic (Weekley & Jones, 1997). Several benefits of video-based SJTs stem from this. First, the detail and realism of video-based SJTs may increase the fidelity of the situational judgment test and result in higher predictive validity of the instrument than a paper-and-pencil based equivalent might demonstrate (Weekley & Jones, 1997). Lievens and Sackett (2006) found some evidence of this in a field study. A video-based SJT designed to assess interpersonal and communication skills demonstrated significantly higher validity for predicting medical students' performance in interpersonally-oriented courses than did a written SJT.

A second advantage of video-based SJTs over their written counterparts is their higher realism. The realism video-based SJTs offer allows job applicants to get a better sense of the demands of the job for which they are applying. Applicants may choose to withdraw from the application process if they discover that the job may not be a good match (Jones & DeCotiis, 1986). Realism leads to the third advantage of video-based SJTs: realism is likely to be associated with a high level of face validity, leading applicants to react more favorably to this medium than they might to a written version of the test (Weekley & Jones, 1997). Several studies have found evidence of enhanced favorability in reactions to video-based SJTs. Chan and Schmitt (1997) administered a situational judgment test of work habits and interpersonal skills to undergraduate students in a lab setting. Students took either a paper-and-pencil or video-based

format of the test. The researchers found that students perceived the video-based format to have higher face validity with respect to a specific job they were asked to consider. In a later study, Richman-Hirsch, Olson-Buchanan, and Drasgow (2000) administered one of three different formats of a situational judgment test of conflict resolution to a sample of managers. These formats were paper-and-pencil, computerized page-turner, and video-based (multimedia). They compared reactions to the multimedia version with pooled reactions to the other two formats and found that managers perceived the multimedia version to have higher content and predictive validity. Managers also found the multimedia version more enjoyable and were more satisfied with the overall assessment process when completing that version of the test.

Meta-analytic findings show that White test takers score higher on average on SJTs than do Black, Hispanic, and Asian examinees (Whetzel, McDaniel, & Nguyen, 2008). However, video-based SJTs may exhibit lower adverse impact than their written counterparts (Lievens & Sackett, 2006). Chan and Schmitt (1997) found that a video-based SJT demonstrated less adverse impact against Black students than did a paper-and-pencil version of the test. Black and White students' scores on the written SJT were positively correlated with their scores on a test of reading comprehension. Scores on the video-based SJT, however, were nearly unrelated to reading comprehension scores, suggesting that the larger adverse impact seen on the written SJT may have been due in part to its reading comprehension component. Jones and DeCotiis (1986) make a related point about an advantage of video-based SJTs: individuals who may perform poorly on written tests, due perhaps to English not being their native language, may perform relatively better on video-based tests because they can see and hear the relevant information for each item.

In research that makes use of both written and video-based versions of an SJT in studying outcomes of interest, the typical approach is to begin with a video-based test and create a paper-and-pencil version of it (see Lievens & Sackett, 2006 and Chan & Schmitt, 1997), rather than vice versa. This makes sense given that researchers tend to acquire these expensive video-based instruments from organizations rather than develop them on their own. Researchers creating paper-and-pencil versions of existing video-based SJTs must make important decisions regarding the amount of contextual information to transfer to the written test. As an example, Lievens and Sackett (2006) developed a written SJT by simply transcribing the dialogue from the video-based SJT. They excluded description of the non-verbal aspects of the video-based test from the written version. This choice seems consistent with Weekley and Jones's (1997) suggestion that video-based SJTs may introduce *irrelevant* contextual information and error into a test. There is little consideration of the implications of such decisions in the research literature on SJTs.

By virtue of their medium, developers of video-based SJTs have to decide not only what kind of contextual information to include but also who to cast as actors in the various scenarios composing the test. Test developers might make the strategic choice to include actors of various demographic backgrounds so as to signal to applicants that the hiring organization welcomes diversity (Avery & McKay, 2006), but to the best of my knowledge, there is no research with SJTs that examines what implications these casting choices have for either the test performance or reactions of applicants from different demographic groups. Research has yet to show whether carefully considering casting choices is imperative or of little import. For video-based SJTs that test interpersonal skills and therefore present test takers with scenarios where two or more individuals are experiencing an interpersonal dilemma, one might expect that responses test takers give to these situations will be influenced by actors' demographic characteristics to the

extent that individuals' evaluations of and consequently behavior toward others is affected by observed demographic characteristics. Theories and research findings in psychology, sociology, and marketing can be brought to bear on this question.

Relational Demography

Relational demography is the term used to describe the comparison of the demographic profiles of two or more individuals who interact with one another (Tsui, Egan, & O'Reilly, 1992). Research on relational demography in the workplace focuses on implications that the demographic similarity between two or more individuals has for the quality of their interactions with one another and for their individual work outcomes. Much of relational demography research is based on the similarity-attraction paradigm (Byrne, 1971 as cited in Lau, Lam, & Salamon, 2008), which proposes that individuals tend to be attracted to those who are similar to them. Identifying with similar others reinforces one's own identity (Steele, 1988) and has been referred to as the self-continuity principle (e.g., Goldberg, Riordan, & Zhang, 2008).

The similarity-attraction paradigm considers the main source of interpersonal attraction to be similarity in attitudes and research evidence indicates that attitude similarity does lead to attraction (Huston & Levinger, 1978). Individuals can infer similarity in attitudes between themselves and others by using demographic cues such as race or sex (Tsui et al., 1992). Indeed, research has found race and sex similarity to be important factors in interpersonal attraction (Ibarra, 1992; Thomas, 1990); individuals tend to perceive those demographically similar to them more positively (Chatman & Flynn, 2001; Tsui & O'Reilly, 1989; Wesolowski & Mossholder, 1997). Graves and Elsass (2005) cite research to suggest that individuals may have more positive expectations about how similar others will behave and treat these individuals accordingly.

Given its influence on interpersonal attraction, demographic similarity is expected to result in a range of positive outcomes for interacting parties. Demographic *dis*similarity, on the other hand, is expected to result in more negative outcomes. Support is a frequent outcome of interest. Goldberg, Riordan, and Zhang (2008) showed that White and Black employees may see their supervisors as better role models when those supervisors are of the same race as them. Similarly, employees may receive more instrumental support (i.e., support, such as challenging assignments and sponsorship, that advances one's career) from their assigned mentors when the protégé and mentor are of the same race (Ensher & Murphy, 1997). Another study showed that employees may perceive a higher level of family supportive supervision when their supervisors are of the same race and/or sex as them relative to when supervisors are of a different race and sex (Foley, Linnehan, Greenhaus, & Weer, 2006).

Performance ratings are another frequently considered outcome in this stream of research. Tsui and O'Reilly (1989) found that supervisors may like their dissimilar subordinates less than similar ones and give dissimilar subordinates lower performance ratings. The tendency for White and Black superiors to give their dissimilar subordinates lower ratings has also been found in the context of developmental performance ratings for managers (Mount et al., 1998). A meta-analysis of race differences in ratings of performance showed that White and Black raters tend to give higher ratings to individuals of the same race as them (Kraiger & Ford, 1985). Similarly, in the context of interview ratings, assessors have been shown to evaluate candidates of a different race less favorably than candidates of their own race (Buckley et al., 2007). Additional evidence of demographic similarity effects comes from cross-sectional survey research in the marketing field, which considers the implications of demographic differences between communicators featured in advertisements and viewers of these ads. This research indicates that Black and White

individuals tend to prefer ads that feature spokespeople of the same race as them (see Whittler, 1991 for a review). Findings of sex similarity effects are infrequently mentioned in the above review because considerably less evidence of these effects has accrued relative to evidence of race similarity effects. There has been investigation of moderators in research on sex similarity effects, as well as in research on race similarity effects. Moderators help to elucidate conflicting findings in relational demography research and are reviewed next.

Moderators of demographic similarity effects. While research has shown demographic similarity to be related to positive outcomes in a variety of contexts, there has also been evidence of asymmetrical effects of demographic similarity across demographic groups. Race as a moderator is discussed first, and is followed by discussion of sex as a moderator. Finally, a rationale for asymmetrical effects is provided.

Race similarity effects may be stronger for White than for minority (e.g., Black) individuals. Goldberg (2005) found, for instance, that in the context of applicant selection, White recruiters showed a stronger preference for applicants of their own race than did Black recruiters. Reviews of rater-ratee race effects on performance ratings (Oppler et al., 1992; Pulakos et al., 1989; Sackett & DuBois, 1991) following those of Kraiger and Ford (1985) cited earlier, reported findings inconsistent with those of Kraiger and Ford for Black raters. While Whites were again found to rate those of their own race higher, Blacks did not rate Blacks higher on the vast majority of performance dimensions. In fact, Blacks may actually rate Whites slightly better than they rate those of their own race (Sackett & DuBois, 1991). Finally, research into the effects of race composition of work groups suggests that for Whites increasing difference from others in the work group is associated with lower organizational attachment, while for nonwhites, being different in race is not related to attachment (Tsui et al., 1992).

Evidence for asymmetrical sex similarity effects is more mixed and comes mainly from investigations of work groups (as opposed to dyads). Tsui, Egan, and O'Reilly (1992) found that being different from those in one's group in sex has a more negative effect on men's psychological attachment to the group than on women's. While increased differences in the sex composition of the group were related to lower attachment for men, women responded to increased differences in sex composition of the group with higher levels of attachment. Chatman and O'Reilly's (2004) findings were consistent with Tsui and colleagues': men were more likely to indicate a preference for leaving female dominated groups than were women to indicate a preference for leaving male dominated groups. On the other hand, Kirchmeyer (1995), who conducted a longitudinal study of the effects of demographic similarity to one's workgroup on outcomes (e.g., work group fit, supervisor support, organizational commitment, turnover, and promotion), did not find sex to moderate the influence of demographic similarity on outcomes; dissimilarity did not have a more negative influence on men than on women. In the case of promotion as an outcome there was actually an unexpected finding in that men benefited from being dissimilar (in sex) from the work group. Finally, Hinds and colleagues (2000) asked students to designate individuals they would like to work with and did not find individuals to show a preference for working with men. (In fact, they did not even find sex similarity effects; sex similarity did not play a significant role in group member choice).

Asymmetrical effects of demographic similarity have been explained in terms of groups' differences in social status. Chattopadhyay, Tluchowska, and George (2004) propose that members of low status groups (e.g., females, minorities) may benefit less than members of high status groups (e.g., males, Whites; Elsass & Graves, 1997) from identifying with other members of their demographic groups because identifying with those of a low status does not enhance

one's identity. Members of low status groups can enhance their identities by associating, instead, with those in the out-group. Thus, unlike individuals of high status, who maintain a positive identity by evaluating members of their out-group less favorably than members of their in-group (Goldberg, 2005), those of low status may not view their out-group more negatively than their in-group. It is important to note, however, that there may be individual differences in how highly identified members of low status groups are with their in-group. Both the relational demography and marketing literatures indicate that ethnic identification may moderate demographic similarity effects. Experimental research studies indicate that preference for one's own group is stronger among those who are highly identified with their group (Whittler, 1989; Williams & Qualls, 1989).

Before wrapping up the discussion of moderators it is important to note that there is some evidence in the marketing literature to show that the effects of demographic (dis)similarity for high status group members are not always straight forward either. Attitudes toward other racial groups appear to moderate the effects of race similarity on White individuals. Whittler and DiMeo (1991) found that White individuals with high scores on a measure of prejudice reported feeling stronger identification with and similarity (in terms of lifestyle, background, appearance, dress, and values) to White actors than Black actors; Whites who received low scores on the measure of prejudice identified equally with and saw themselves as equally similar to White and Black actors.

Time frame for demographic similarity effects. The very short time frame over which test takers view video-based SJT items might appear to pose a problem when trying to draw on relational demography research to make predictions about the role demographic similarity will play in respondents' evaluations of (and consequently behavioral intentions toward) actors in an

SJT. On the contrary, demographic similarity effects will be especially relevant with the short time. It is important to note that the effects of demographic (dis)similarity are expected to be fairly immediate (and to occur even absent interaction between individuals; Tsui et al., 1992) and to actually decrease with time (Watson, Kumar, & Michaelsen, 1993). The expected immediacy of the effects is supported by models of impression formation, which suggest that individuals process others' social category (including demographic) information automatically (Brewer & Harasty Feinstein, 1999; Fiske, Lin, & Neuberg, 1999). More specifically, research shows that when one encounters a member of a certain group about whom he or she has no individuating information—that is information specific to that individual as opposed to information about the individual's social category (inferred from characteristics such as race, sex, and age)—his or her judgment of that group member will be influenced by that individual's social category (Kunda & Thagard, 1996 provide a discussion of this research). The effects of demographic (dis)similarity should decrease with time, as Watson and colleagues (1993) suggest, because with time, individuating information about individuals becomes available. Thus, one might expect the types of effects that have been found in relational demography research to generalize to the context of items on a video-based SJT, where respondents have little more than the actors' demographic information upon which to base their evaluations of these individuals.

Demographic similarity and behavioral intentions. Respondents' evaluations (affective, cognitive) of the actors in the videos, based on the demographic similarity between the respondent and the target actor, can be expected to influence respondents' behavioral intentions. This expectation is consistent with research linking evaluations of targets with behavior toward those targets. For example, teachers' interactions with students who they perceive to be more (less) capable tend to be characterized by better (worse) quantity and quality

(see Reyna, 2008 for a review of the consequences of stereotypes in classroom settings). In the context of helping situations, evaluations of help seekers (how responsible are they for their problems?) are often found to be related to helpers' behavior and intentions (MacGeorge, 2003).

To summarize, the relational demography perspective, if applied to video-based SJTs, would suggest that demographic (dis)similarities will have effects on respondents' evaluations of target actors. Evaluations of target actors, in turn, will influence respondents' behavioral intentions. Notably, particular types of situations may prove important in the operation of demographic similarity effects. Given that individuals are socialized to behave in certain ways in particular situations (Wilson, Lizzio, Whicker, Gallois, & Price, 2003), different situations may elicit different patterns of behavior, and potentially temper relational demography effects on evaluations and behavioral intentions. Two particular contexts that have received attention in the social categorization literature are helping situations and conflict situations. These situations are also commonly found in SJTs. In the next section helping situations will be reviewed and hypotheses for helping-type SJT scenarios will be presented. Afterward, conflict situations will be reviewed and hypotheses for conflict-type SJT scenarios will be proffered.

Helping Situations

Helping behavior is considered a form of citizenship behavior ("performance that supports the social and psychological environment in which task performance takes place"; Organ, 1997, p. 95) and involves helping coworkers solve (or prevent) work-related problems (Podsakoff, MacKenzie, Paine, & Bachrach, 2000). A meta-analysis conducted by Podsakoff, Whiting, Podsakoff, and Blume (2009) found that organizational citizenship behaviors are related to both individual-level (e.g., managerial ratings of employee performance, absenteeism, turnover intent and turnover) and organizational-level (e.g., productivity, customer satisfaction,

unit-level turnover) outcomes. Earlier, Podsakoff and colleagues' (2000) narrative review of the literature tied helping behaviors, in particular, to indicators of organizational-level outcomes (e.g., efficiency, performance quantity and quality, customer satisfaction). Given these relationships between helping behavior (and organizational citizenship behavior in general) and outcomes organizations value, Podsakoff and colleagues (2009) recommended that organizations select individuals who will be likely to perform these behaviors. It is not surprising then, that SJTs often include help seeking (or giving) situations to determine applicants' willingness to offer help to coworkers. Research on helping was consulted to determine the role that demographic similarity may play in individuals' propensity to provide help.

Helping and race similarity. Milliken and Martins (1996) observed that when differences (e.g., racial, gender) between people can be clearly seen, biases, prejudices, and stereotypes can cause individuals to respond to these differences in particular ways. Helping researchers have been particularly interested in how individuals respond to observed differences in race between themselves and those that they are in a position to help. Helping of Blacks (relative to Whites), who have historically been the targets of negative attitudes and discrimination in our society, is most often of interest in this research. Saucier, Miller, and Doucet (2005) conducted a meta-analysis of studies looking at differences in White individuals' helping of other Whites and Blacks. While they did not find evidence of universal discrimination against Blacks, they did find that Whites helped Blacks less than they helped those of their own race in the same situations when they had more ways to rationalize withholding help without needing to implicate race. In other words, when the situation provided nonracist justification for not helping a Black individual Whites helped Blacks less. Help given to Blacks decreased as compared to help given to Whites as the following factors increased: time it took to help, risk for

the helper, difficulty of helping, and effort of helping. Findings were mostly consistent with the theory of aversive racism (e.g., Hodson, Dovidio, & Gaertner, 2002), which proposes that while consciously believing in equality between the races and in their own lack of racist attitudes, Whites could experience discomfort (e.g., fear, nervousness) in the presence of Black individuals. Whites are not expected to act on this discomfort and show hostility or obvious discrimination because of the cognitive dissonance (i.e., lack of consistency between one's professed attitudes and one's behavior) this would cause, but they may behave in ways that do not appear obviously racist (e.g., avoid contact). Avoidance behavior will not appear to the non-helper or outsiders as obviously racist when the motivation for it is ambiguous (i.e., can be explained by other situational factors) (Saucier, Miller, & Doucet, 2005).

In a series of experiments, Kuntsman and Plant (2008) found that White participants perceived an emergency situation to be less severe and felt less inclined to provide help when the victim was Black as compared to when the emergency was identical but featured a White victim. Drawing on the aversive racism perspective, the researchers explained Whites' distorted perceptions of the emergency when the victim was Black, in terms of White's desire to both avoid interracial contact with an individual of a race that elicits negative associations (e.g., Black seen as bad) and avoid appearing biased against Blacks. Importantly, the researchers also found that Black participants, unlike Whites participants, did not differ in their perceptions of the emergency depending on the race (Black, White) of the victim, and were equally willing to help the Black and White victims. This finding makes sense in light of research discussed earlier which suggests that demographic similarity may not influence all groups of individuals in the same way (Riordan & Shore, 1997).

Based on relational demography research and Saucier, Miller, and Doucet's (2005) meta-analysis, one might expect that for situational judgment items with scenarios related to helping, the level of help respondents provide will be influenced by whether the help-seeking actor in the scenario ("non-focal actor") is of the same or a different race than the respondent. As research on racism usually focuses on Blacks (Alvarez, Juang, & Liang, 2006), little has been done to examine how race factors into help received by individuals of Asian backgrounds. Moreover, I could find no research that has placed Asians into roles of helpers to examine their tendencies to offer or withhold help to individuals of varying races. Therefore, findings from relational demography research are used and research with White and Black individuals is extrapolated from to make predictions about the effects of race mismatch between the respondent and the non-focal actor for Asian in addition to White and Black individuals.

Extrapolating from Saucier and colleagues' (2005) finding that White individuals helped Blacks less when the situation provided non-race-related reasons for withholding help, I expect that White respondents will offer less help to a minority non-focal actor than to a White non-focal actor. This is likely to be the case because respondents will be able to use situational, race-unrelated cues in the test items to justify offering a low level of help to the minority non-focal actor or withholding help altogether. Based on research showing asymmetrical effects of demographic similarity for high and low status groups (probably because identifying with a higher status group rather than with their own low status group is reinforcing for low status individuals), one might expect that a racial mismatch between the respondent and the non-focal actor will have a more negative influence on help offered by White as compared to minority respondents.

Hypothesis 1: Respondent race will moderate the effects of the race mismatch between the respondent and the non-focal actor on helping, such that the effects of race mismatch will be more negative for White respondents viewing a minority non-focal actor than for minority respondents viewing a White non-focal actor.

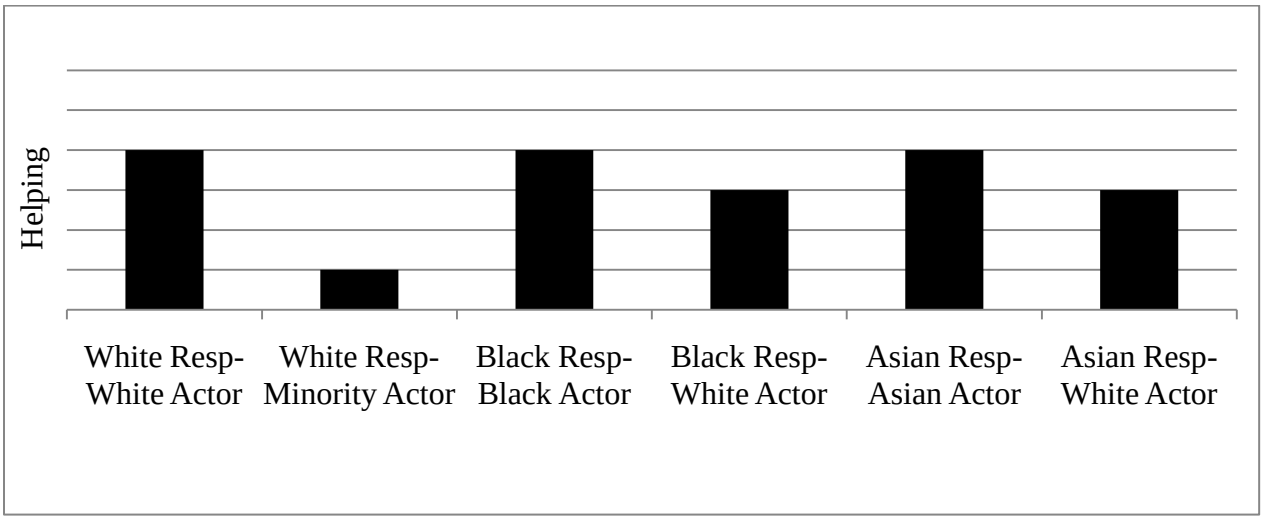

Figure 1. **Illustration of Hypothesis 1.**

Helping and sex similarity. A considerable amount of research has investigated gender effects on helping. In her review of the literature, Eagly (2009) notes that when it comes to organizational citizenship behaviors, women (as compared to men) seem to perform more behaviors ("altruism") that are targeted toward specific individuals who need assistance, while men (as compared to women) appear to perform more behaviors ("civic virtue") that show extra commitment to the organization (versus being aimed at particular coworkers). The former type of citizenship behavior is of more relevance to the type of helping SJT items in the present situation. I expect, therefore, that on average, female respondents will offer more help to non-focal actors than will male respondents.

Hypothesis 2: There will be a main effect for respondent sex on helping, such that female respondents will offer more help to non-focal actors than will male respondents.

Interestingly, meta-analytic evidence on the role of gender in helping (i.e., Eagly & Crowley, 1986) does not support the notion of sex similarity effects in this context. Eagly and Crowley (1986) concluded that men are significantly more likely to help women than they are to help other men, and women are about as likely to help men as they are to help other women. What this suggests then is interactive effects of target sex and respondent sex on helping. These interactive effects may be related both to the consistency (or lack thereof) of targets' behavior with gender roles as well as men's benevolent sexism toward women.

Behavior consistent with gender roles (commonly held beliefs about male and female individuals; Eagly, 2009) tends to be met with approval whereas behavior at odds with gender roles tends to be punished (e.g., Eagly, Karau, & Makhijani, 1995). Gender norms for men prescribe autonomy and reliance on oneself (Auster & Ohm, 2000) and there is evidence to suggest that when men are seen as departing from these proscriptions (e.g., by asking for help), other men may react more negatively than women. In fact, women have been found to have more egalitarian gender role orientations than men (e.g., Judge & Livingston, 2008). MacGeorge (2003) found that men presented with a male's request for help showed a greater tendency than did women to blame the male (as evidenced by perceiving him as more responsible for his own problems and becoming more angry at him). Women are stereotyped as being weak, needy, whiny, and lower in competence relative to men (Kidder & McLean Parks, 2001). Thus, when women ask for help, they are seen as acting consistently with expectations for their behavior. Furthermore, paternalistic protectiveness (exhibited, perhaps, via politeness and courteous acts)

may be used to reward women whose behavior is appropriately guided by gender roles (Glick & Fiske, 2001). Thus, when women ask for help, male helpers are expected to be more willing to assist them due to their "acting like women" than to assist males due to their being seen as violating gender norms for males. Women, with their more egalitarian views toward gender norms should provide equal assistance to male and female requesters.

Note that given the divergent evidence provided by research on helping and relational demography in regards to gender and gender similarity effects, the relatively high inconclusiveness of relational demography research in regard to these effects, and the higher relevance of research on helping to the present study, more weight is being placed on findings from helping research when forming the following hypothesis:

> Hypothesis 3: Respondent sex will moderate the effects of non-focal actor sex on helping, such that a female non-focal actor will receive more help than a male non-focal actor when the respondent is male. Female and male non-focal actors will receive equal amounts of help when the respondent is female.

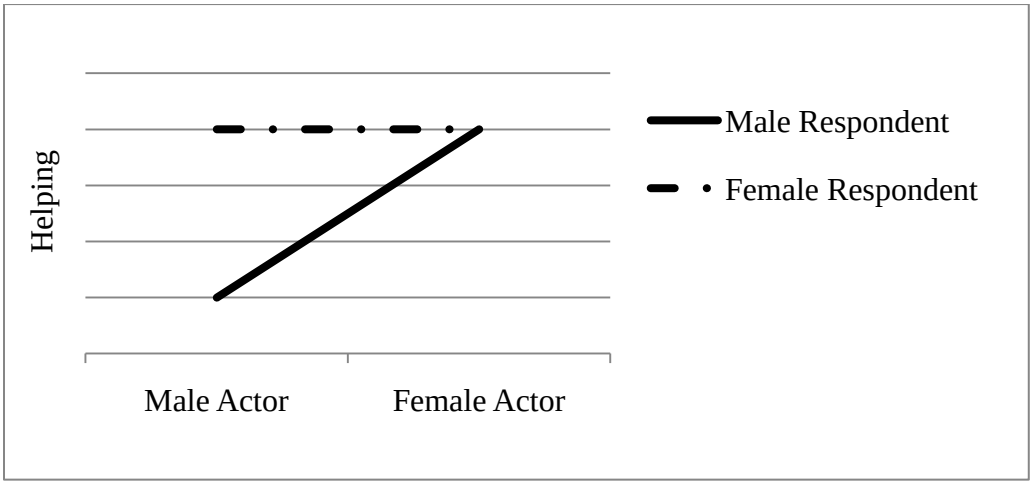

Figure 2. **Illustration of Hypothesis 3.**

Race *and* sex considerations in helping. Because individuals belong to multiple groups (race, gender), one needs to consider which of their identities individuals will use to relate to actors in situational judgment items. Researchers (e.g., Whittler & DiMeo, 1991; Williams & Qualls, 1989) often assume that individuals will primarily use race to identify with others, regardless of whether or not identification based on other characteristics (e.g., gender) is possible in the given situation. This is consistent with the ethnic prominence hypothesis proposed by Levin and colleagues (2002), which suggests that in cases where multiple identities intersect, ethnicity will be most salient. Brumbaugh (2009) questions this assumption and demonstrates that identification with others is not always based mainly on race; it depends on contextual cues which make race more or less prominent in a given situation. Indeed, Browne and Misra (2003) in their review of the literature on the intersection of race and gender point out the inconclusiveness in the findings (will ethnicity [Levin, Sinclair, Veniegas, & Taylor, 2002], gender [Sidanius & Pratto, 1999], or both [Landrine, Klonoff, Alcaraz, Scott, & Wilkins, 1995; Nelson & Probst, 2004] determine outcomes?) and argue for the need to carefully consider context. Brumbaugh's (2009) findings were actually not that surprising given that context has been recognized earlier as influencing the salience of various identities (Ridgeway & Smith-Lovin, 1999 provide a review).

In the context of situations where an individual is asked for assistance, he or she would be likely to consider the competence and work ethic of the individual seeking assistance. However, when no additional information about the target beyond social category is available, stereotypes (beliefs about social groups) will provide information about target individuals for processing (Landy, 2008). When it comes to competence, women, as well as Blacks, are stereotyped as being less competent than men and Whites, respectively. Accordingly, women

and Blacks are expected to need to ask for help. When it comes to work ethic, Blacks are stereotyped as lazy while Asians are stereotyped as hard working (e.g., Reyna, 2008). Therefore, Blacks might be seen as asking for help so as to avoid doing work themselves while Asians might be seen as asking for help because they genuinely need assistance. Timberlake and Estes (2007) demonstrated that stereotypes are amenable to categorization along dimensions of race and sex salience (p.404). Using this scheme, the competence stereotype could be argued to have high race and high sex salience while the work ethic stereotype could be argued to have high race but low sex salience. To the extent that one or both of these stereotypes will be activated in a particular helping situation, the sex, race, or both sex and race of the target will be salient to the respondent.

Considering the two types of helping situations in the present study (these two situations are described in the methods section), the competence stereotype is more likely to be relevant for one situation while the work ethic stereotype is more likely to be relevant for the other. When the competence stereotype is activated, and both race and sex is salient to respondents, the lowest level of help is likely to be offered when a White male respondent is combined with a Black non-focal actor as compared to other combinations of respondents with targets. This expectation is based on the research already discussed. As explained earlier (recall Hypothesis 1), the effects of race mismatch are expected to be more negative for White respondents viewing a Black non-focal actor than for minority respondents viewing a White non-focal actor, and men are expected to help other men less than they help women (recall Hypothesis 3).

Hypothesis 4: In a helping situation where the competence stereotype is activated, the lowest level of help will be offered by White male respondents to male Black non-focal actors as compared to other race and sex-based combinations of respondents with targets.

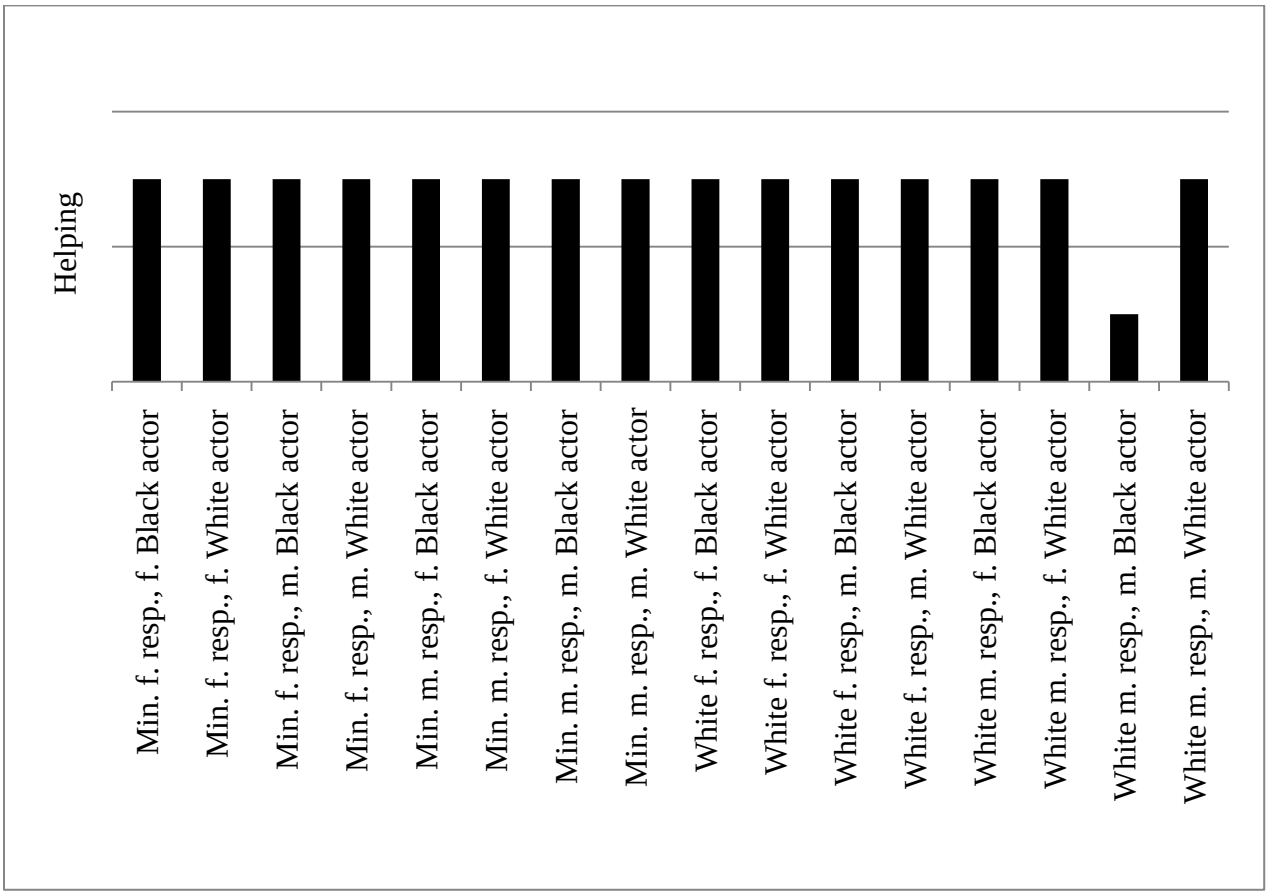

Figure 3. **Illustration of Hypothesis 4.**

When the work ethic stereotype is activated, and race is most salient to respondents, the lowest level of help is likely to be offered when a White respondent is combined with a Black non-focal actor. This expectation is, in effect, subsumed within hypothesis 1, but a separate hypothesis is proposed because one particular type of helping situation is being considered at this time.

Hypothesis 5a: In a helping situation where the work ethic stereotype is activated, the lowest level of help will be offered by White respondents to Black non-focal actors as compared to other race-based combinations of respondents with targets.

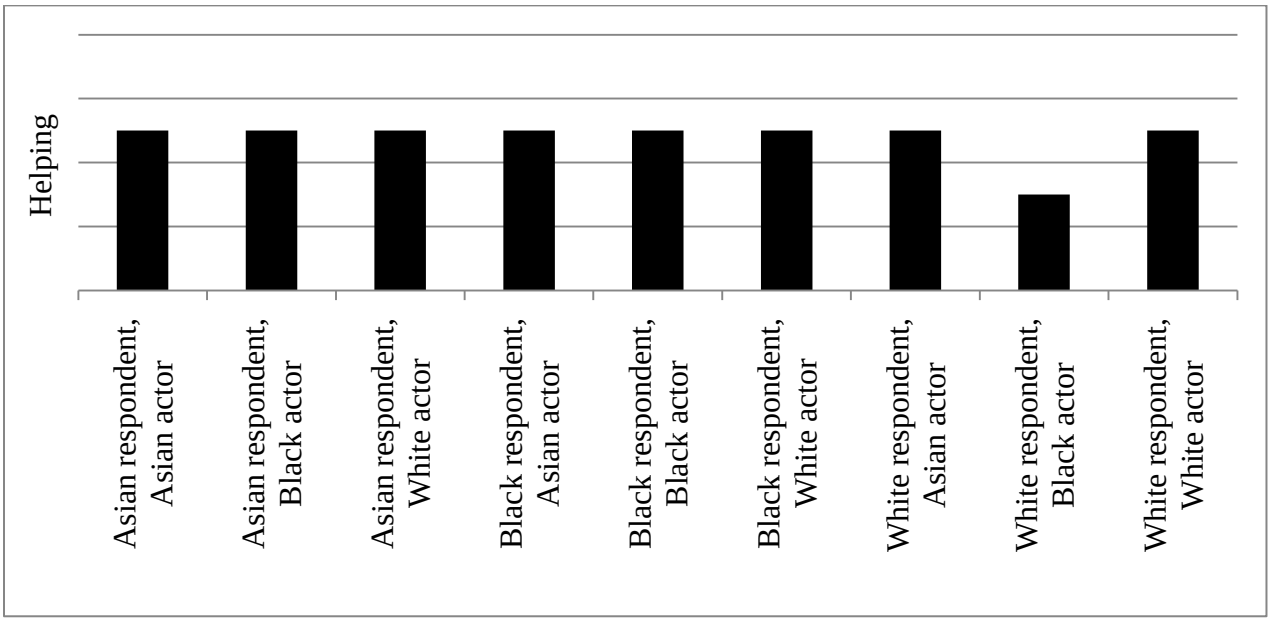

Figure 4. **Illustration of Hypothesis 5a.**

Asians are excluded from being grouped with Blacks in the above hypothesis because of their status as the "model minority" (Alvarez et al., 2006). Recall the stereotype that Asians have high work ethic. When this stereotype is activated, the level of help received by Asian non-focal actors may even exceed the level of help received by White non-focal actors (the high status group). (Notably, if this prediction is supported, there will be some inconsistency with Hypothesis 1).

Hypothesis 5b: In a helping situation where the work ethic stereotype is activated, the highest level of help will be offered to Asian non-focal actors.

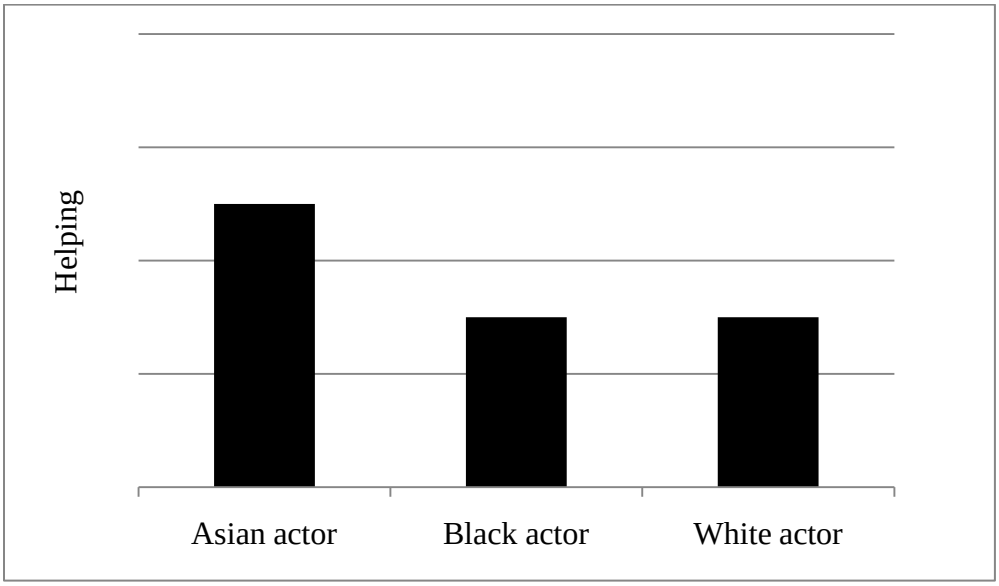

Figure 5. **Illustration of Hypothesis 5b.**

Conflict Situations

The term "conflict" is used to describe a situation where interacting individuals have apparently incompatible concerns (Thomas, Thomas, & Schaubhut, 2008). The success of individual employees and the teams they work within depend in part on individuals' ability to deal effectively with conflict situations that arise (Tjosvold, 1998), with managers, for example, spending 20 percent of their time, on average, dealing with conflict (Thomas, 1992, as cited in De Dreu et al., 2001). As organizations become more diverse as well as more reliant upon teams of individuals to work together effectively, of substantial concern is the potential for conflict between individuals in demographically heterogeneous groups (King, Hebl, & Beal, 2009). Given this concern, it is not surprising that organizations, as part of their selection process, should wish to examine applicants' (particularly for managerial positions) ability to effectively handle conflict situations. Thus, interpersonally-oriented SJTs are likely to incorporate items dealing with conflict situations (e.g., Chan & Schmitt, 2002; Mumford et al., 2008), or to be

composed entirely of conflict situations (e.g., Olson-Buchanan et al., 1998). Research on conflict was consulted to determine the role that race and sex may play in how individuals choose to handle conflict.

Several frameworks describing conflict situations have been proposed. One such framework comes from K. W. Thomas' work. Two dimensions are used for mapping the parties' intentions: cooperativeness and assertiveness (Thomas et al., 2008). Cooperativeness refers to the attempt to satisfy the other party; assertiveness refers to the attempt to satisfy one's own interests. Various combinations of these two dimensions (e.g., low cooperativeness with low assertiveness, high cooperativeness with high assertiveness) constitute the type of conflict resolution strategy an individual employs.

Conflict and race similarity. Not inconsistent with the relational demography perspective, researchers generally expect individuals to exhibit a tendency towards cooperating more with similar others than with dissimilar others (Tajfel, 1978, as cited by King et al., 2009). Consistent with the earlier discussion of the time frame of demographic similarity effects, cooperation is expected to be difficult at early stages of individuals' interactions because based on limited information about one another, individuals categorize each other (as being in the out-group) according to visible demographic characteristics (Chatman & Flynn 2001). In support of this expectation, Chatman and Flynn (2001) found that demographic diversity between individuals within groups is negatively related to perceptions and development of cooperative norms when the group is starting out. With regards to race differences between individuals in particular, it is argued that one's race will influence perceptions and subsequently reactions to others during conflict situations (Davidson, 2001). Specifically, conflicts often start with an episode of potentially provocative behavior by one of the parties, which the other party has to

interpret in order to decide how to respond. The characteristics of the parties, including race, should help to determine what the interpretation will be. Based on ideas from relational demography, one might expect *less* negative interpretations of provocative acts among racially similar (i.e., in-group members) individuals (Davidson, 2001). Surprisingly, though, researchers sometimes find main effects for target's race and the nature of these main effects is not even consistent. Sagar and Schofield (1980), for instance, found that children (both Black and White) judged ambiguous aggressive behavior more negatively when it was done by a Black as opposed to a White individual. In Davidson's (2001) experiment, both Black and White participants interpreted a White manager's provocative act (taking credit for another individual's work) more negatively than a Black manager's. These inconsistent findings may have something to do with differences in situational context (e.g., Blacks are stereotyped as being more aggressive than Whites, Whites are seen as more exploitative than Blacks), reaffirming again the need to consider context in making predictions regarding race effects in various situations.

With regard to actual responses to conflict, research points to some potential differences between racial groups. Davidson (2001) found that Whites prefer avoiding conflict more than Blacks do (Blacks are more comfortable with the idea of being confrontational than are Whites), but both groups endorse an avoidance strategy more so when considering conflict with those of a different race (only Black and White were considered). According to the conflict framework described above, an avoidance strategy is defined as being low on the assertiveness as well as low on the cooperativeness dimensions (Thomas et al., 2008). The expectation that individuals will show higher cooperativeness toward racially similar others is consistent with the relational demography perspective. The expectation that they will show higher assertiveness toward racially similar others may seem inconsistent with the relational demography perspective but

perhaps it is not. Just as one might argue that liking another individual more (due to similarity with him or her) should make someone less assertive with that individual, one can also argue that feeling more comfortable in an interaction with another individual (again, due to similarity with him or her) should make someone more willing to be assertive (i.e., interpersonal comfort may underlie willingness to be assertive). To the extent that findings from Davidson's (2001) study generalize, individuals may potentially show lower assertiveness and lower cooperativeness when responding to conflict with an individual from a different racial group.

Hypothesis 6a: In responses to conflict situations there will be a main effect for respondent-non-focal actor race match on assertiveness such that respondents will endorse a more assertive response when the respondent and the non-focal actor are of the same race than when they are of different races.

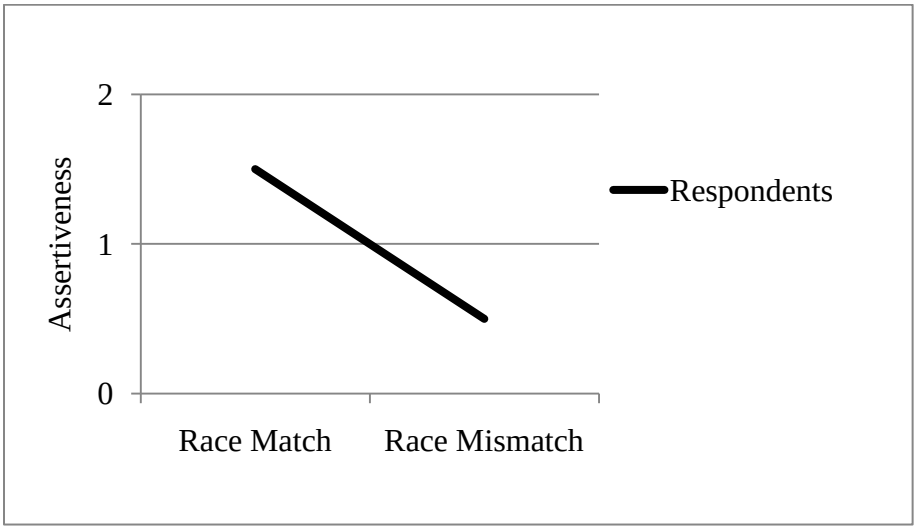

Figure 6. **Illustration of Hypothesis 6a.**

Hypothesis 6b: In responses to conflict situations there will be a main effect for respondent-non-focal actor race match on cooperativeness such that respondents will

endorse a more cooperative response when the respondent and the non-focal actor are of the same race than when they are of different races.

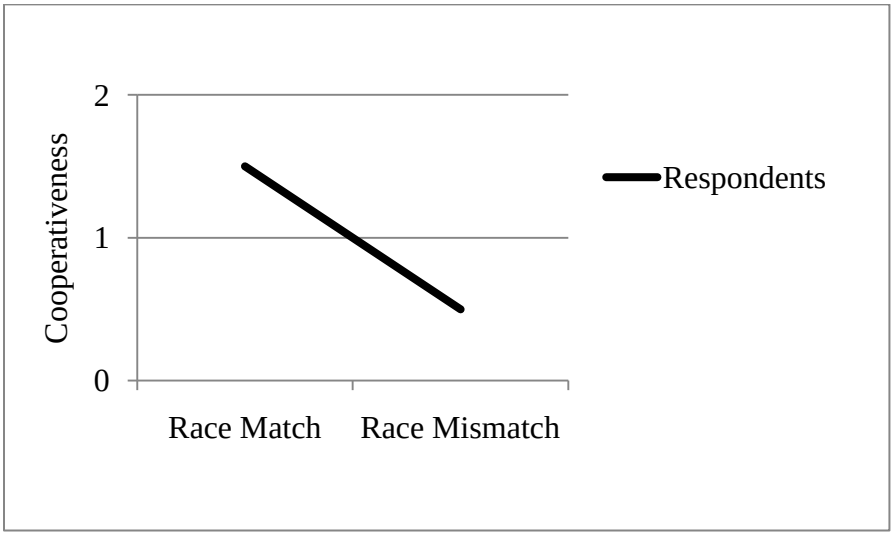

Figure 7. **Illustration of Hypothesis 6b.**

Another consideration in conflict situations is whether or not a third party is brought in to try to resolve the conflict. The decision to involve a third party may arguably be related to cultural differences between individuals. Those who are collectivist (i.e., emphasis on the interdependence of group members) may conceivably be more likely to seek the help of a third party than those who are individualistic (i.e., emphasis on independence from group members and self-reliance) because of the desire to avoid confrontation and possible damage to the relationship with the other party (Giebels & Yang, 2009). While there is no complete consensus on this being the case (see Oyserman, Coon, & Kemmelmeier, 2002), research suggesting that Asians, Hispanics, and Blacks have more collectivist (defined in that research as orientation toward the welfare of one's larger community) tendencies than do Whites (Gaines et al., 1997) has lead researchers to test for main effects of culture on preferences for help from third parties. Davidson (2001) evaluated but did not find support for the prediction that Blacks would use

social support in conflict situations more so that would Whites. Similarly, Giebels and Yang (2009) did not find differences in general preference for third party intervention between the groups they classified as collectivist (Chinese individuals) and individualist (Caucasian individuals). It may be the case, however, that collectivist groups' preference for third party intervention also depends on whether or not the other party is of the same racial background. That is, collectivist individuals may prefer to get assistance from a third party in conflict situations with members of their group more so than in conflict situations with members of an out-group because maintaining relationships with those in the in-group is perceived as more important. Some support for this position comes from research suggesting that Chinese individuals will be more likely to adopt a competitive approach over a cooperative approach *unless* they are dealing with in-group members with whom the importance of a long-term relationship outweighs the importance of short-term gain (Koch & Koch, 2007 provide a critical analysis of relevant research). Race match, then, might have a positive effect on collectivist individuals' use of third party intervention. There is little evidence to suggest, however, that race match should have an influence on individualist individuals' use of third party intervention. (Note that this hypothesis will apply only when intervention has a reconciliatory goal. This point is further elaborated on below.)

> Hypothesis 7: There will be an interaction between race and respondent-non-focal actor race match on preference for third party intervention when the goal is reconciliation. Respondents belonging to groups classified as collectivist (Asians, Blacks) will show a preference for third party intervention when paired with non-focal actors of the same race as them as compared to when they are paired with non-focal actors of a different race.

Race match between respondents and non-focal actors will not influence preferences for third party intervention among respondents classified as individualist (Whites).

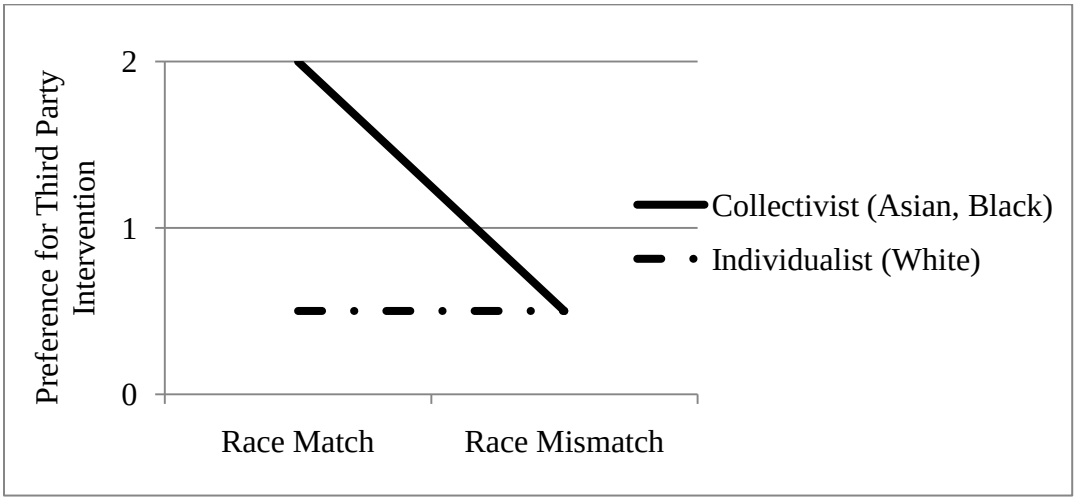

Figure 8. **Illustration of Hypothesis 7.**

The above hypothesis is likely to only hold true for one of the conflict situations in the proposed study, not both. This is due to a divergence in the goals of third party intervention in the two SJT scenarios. While in one (for which the above hypothesis is relevant) intervention appears to have a reconciliatory goal, in the other intervention has more of a report-coworker's-poor-performance goal. The ideas that informed the above hypothesis (i.e., different groups' differential desire to maintain good interpersonal relationships) logically lead to different predictions depending on the goal of the third party intervention. When the coworker is of the same race as them , members of collectivist groups should be *more* likely to prefer an intervention when the goal is reconciliation, but *less* likely to prefer an intervention when the goal is reporting the coworker's poor performance. Reconciliation with a coworker helps to maintain a relationship; getting a coworker into trouble is likely to harm the relationship. As for the case when reconciliation is the goal of the intervention, there is little evidence to suggest that

race match should have an influence on individualist individuals' use of third party intervention when reporting a coworker's poor performance is the goal. Thus, the following is expected for a scenario where third party intervention has the goal of reporting a coworker's poor performance:

> Hypothesis 8: There will be an interaction between race and respondent-non-focal actor race match on preference for third party intervention when the goal is reporting a coworker's poor performance. Respondents belonging to groups classified as collectivist (Asians, Blacks) will show a preference for third party intervention when paired with non-focal actors of a different race as them as compared to when they are paired with non-focal actors of the same race. Race match between respondents and non-focal actors will not influence preferences for third party intervention among respondents classified as individualist (Whites).

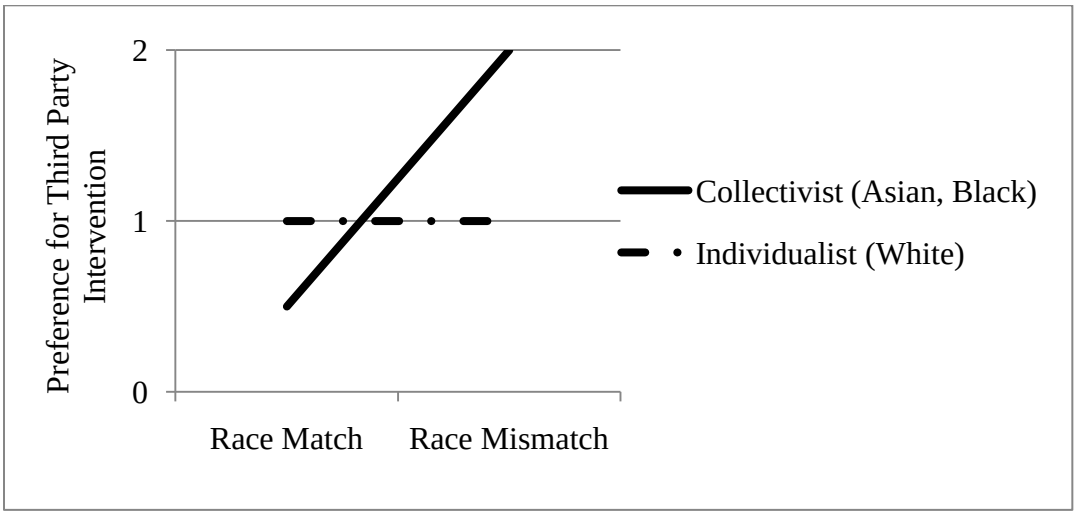

Figure 9. **Illustration of Hypothesis 8.**

Conflict and sex similarity. As previously noted, gender norms prescribe and proscribe certain behaviors for men and women. Women are expected to be more nurturing and

subordinate, while men are expected to be more assertive and dominant (Graves & Elsass, 2005). Accordingly, women are seen as more oriented than men toward maintaining relationships (Wilson et al., 2003). Asserting one's rights is perceived as detrimental to interpersonal relationships and women who behave this way are seen as violating gender norms for women (Wilson et al., 2003). Furthermore, research suggests there is a high level of agreement among people regarding what is socially appropriate behavior for their own and opposite gender groups in particular contexts (e.g., Wilson et al., 2003). For instance, Wilson and colleagues (2003) found that when unfairly criticized, men agree that it is socially appropriate for them to be proactive and direct in asserting that they have been evaluated unfairly. Women, on the other hand, agree that it is fitting for them to be less direct and to allow the criticizer a chance to explain his or her point of view. Women might compromise and accept some amount of responsibility in a conflict situation so as to avoid damaging the relationship with the other party (Wilson et al., 2003). This sort of behavior is, again, in line with gender role expectations for women.

Given that women tend to endorse gender norms which dictate that women should use more affiliative strategies in resolving interpersonal challenges and that men tend to endorse norms that dictate that men are allowed to behave more assertively, female test takers are expected to respond to conflict situations in a more cooperative and less assertive manner than male test takers.

Hypothesis 9a: There will be a main effect for respondent gender on assertiveness when responding to conflict, such that female respondents will respond to conflict with lower assertiveness than will male respondents.

Hypothesis 9b: There will be a main effect for respondent gender on cooperativeness when responding to conflict, such that female respondents will respond to conflict with higher cooperativeness than will male respondents.

Behavior and evaluations of its effectiveness has been shown to be related to the relative gender of the interacting parties (Carli, 1989). In an investigation of gender differences in social influence within dyads, Carli (1989) found that both men and women agreed more and exhibited more positive social behaviors (e.g., displays of positive affect, displays of solidarity) when paired with a female interaction partner. Boys are reported to employ verbal, physical, and indirect (e.g., shut out of group, ignore) aggressiveness less with girls than with other boys (Russell & Owens, 1999). The same study also found girls to be more physically aggressive toward boys than toward other girls. Both male and female adolescents use compromise as a conflict resolution strategy more with females than with males and males display anger less with females than they do with other males (girls, however, display equal amounts of anger with both gender groups) (Shute & Charlton, 2006). Individuals are more accepting of "masculine" behavior directed at men when discussing controversial topics (Carli, 1989) and perceive it as appropriate to be more assertive in response to a man offering criticism than in response to a woman offering criticism (Wilson et al., 2003, Experiment 2).Thus, a considerable amount of research suggests that both men and women seem to perceive it more acceptable to behave in a more masculine way toward men than toward women. These findings apparently reflect attitudes of benevolent sexism toward women (Wilson & Gallois, 1985) and are not inconsistent with findings from the research on helping reviewed earlier.

Given the research that suggests a general tendency to behave more positively toward women (Carli, 1989) as well as the more specific tendency to be less assertive when responding to a criticizing woman as compared to a criticizing man, asymmetrical effects of sex similarity between respondents and non-focal actors in conflict-related SJT items are expected. Men are expected to respond to male actors in a less affiliative manner (higher assertiveness, lower cooperativeness) than to female actors while women are expected to respond to female actors in a more affiliative manner (lower assertiveness, higher cooperativeness) than to male actors.

Hypothesis 10a: Respondent gender will moderate the effects of gender match between the respondent and the non-focal actor on assertiveness when responding to conflict. For male respondents, gender match will have a positive effect on assertiveness. For female respondents, gender match will have a negative effect on assertiveness.

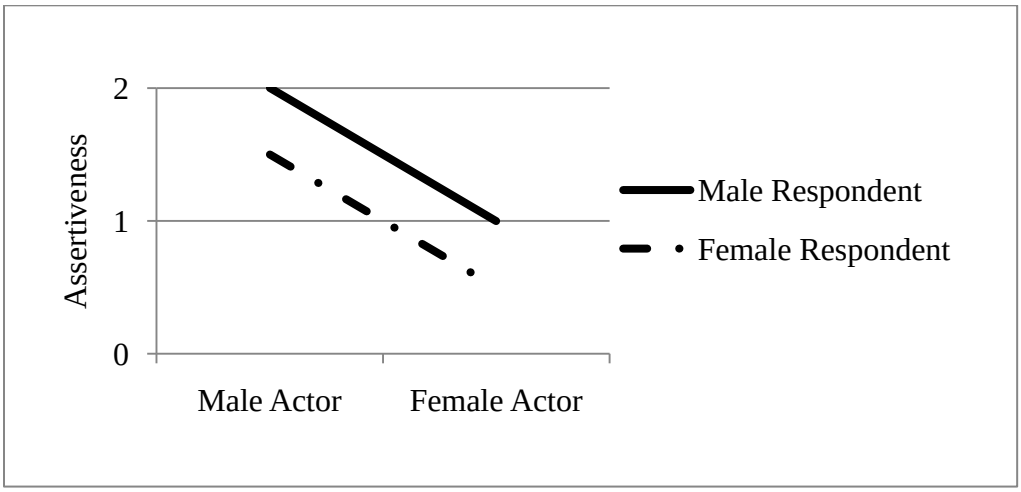

Figure 10. **Illustration of Hypothesis 10a.**

Hypothesis 10b: Respondent gender will moderate the effects of gender match between the respondent and the non-focal actor on cooperativeness when responding to conflict.

For male respondents, gender match will have a negative effect on cooperativeness. For female respondents, gender match will have a positive effect on cooperativeness.

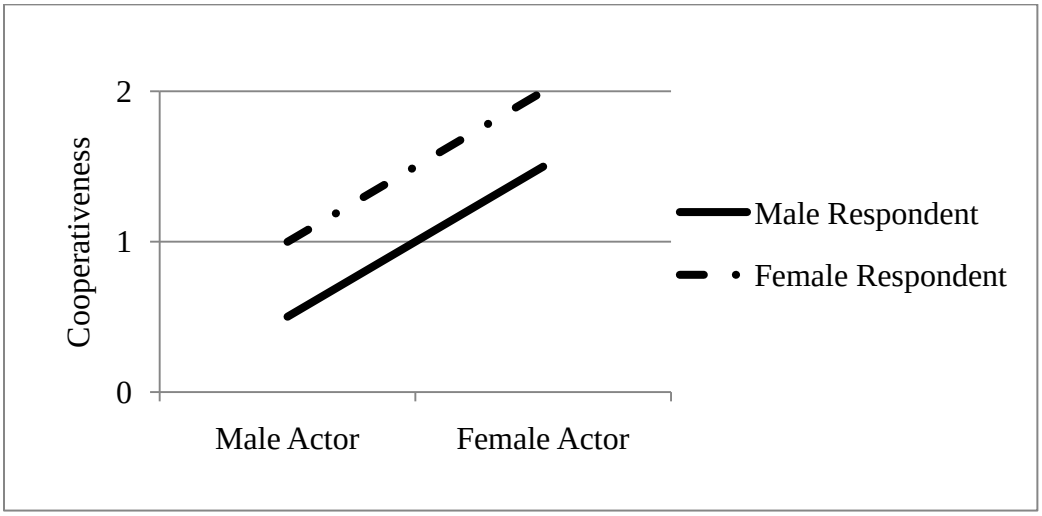

Figure 11. **Illustration of Hypothesis 10b.**

It is necessary to consider the role of gender in determining preferences for third party interventions in conflict situations. Giebels and Yang (2009) did not find differences in general preference for third party intervention between men and women. Interactive effects may be possible, however. As discussed earlier, women are expected to be more nurturing than men (Graves & Elsass, 2005) and more oriented toward maintaining relationships in general and during conflict situations specifically (Wilson et al., 2003). Given evidence that women will tend to endorse such gender norms (e.g., Wilson et al., 2003), a similar process may operate for women wanting to preserve relationships in the face of conflict as was described earlier for collectivist groups. Combined with the expectation that women will respond in a more affiliative manner toward other women than toward men (recall Hypothesis 10), this yields the possibility of a respondent gender by respondent-target gender match interaction on third party intervention preferences. The nature of the interaction should again (recall the difference between hypotheses

7 and 8) be different depending on the type of conflict scenario (i.e., goal of reconciliation, goal of reporting a coworker's poor performance).

The effect of gender match on male respondents is somewhat harder to predict than the effect of gender match for female respondents. Given that men are expected to respond to women in a more affiliative manner than to other men (recall Hypothesis 10), like with female respondents, their preferences may also be influenced by their gender (mis)match with the non-focal actor. Alternatively, however, men may not want third party intervention regardless of the gender of the non-focal actor because it may be seen as less than masculine to seek help in resolving one's interpersonal issues. I tentatively hypothesize that the first alternative will hold true. Thus, it is expected that:

> Hypothesis 11: When the goal of the conflict situation is reconciliation, there will be an interaction between respondent gender and respondent-non-focal actor gender match on preference for third party intervention of the following nature: a. Female respondents will show a preference for third party intervention when paired with female non-focal actors as compared to when they are paired with male non-focal actors; b. Male respondents will show a preference for third party intervention when paired with female non-focal actors as compared to when they are paired with male non-focal actors.

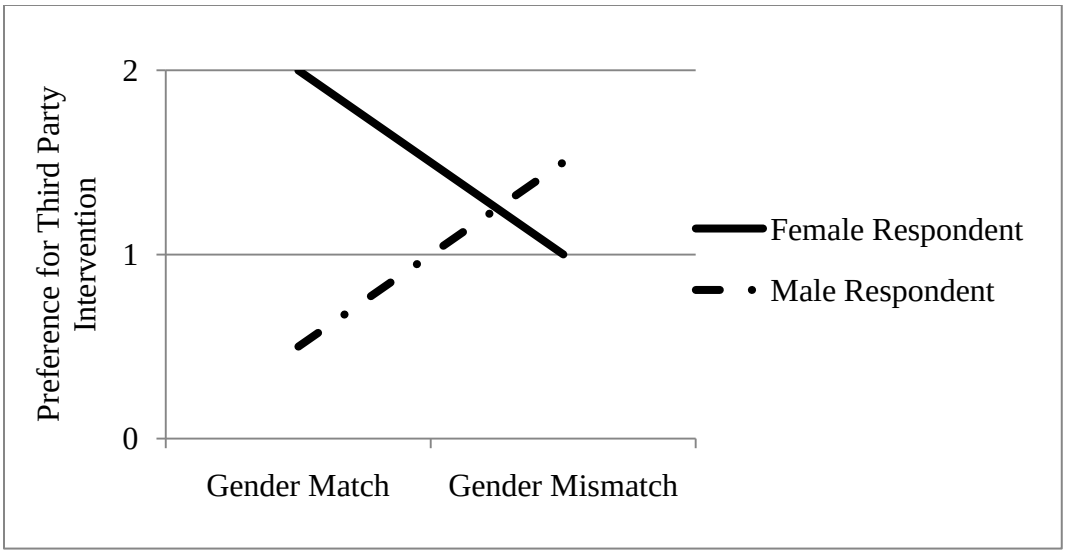

Figure 12. **Illustration of Hypothesis 11.**

Hypothesis 12: When the goal of the conflict situation is reporting a coworker's poor performance, there will be an interaction between respondent gender and respondent-non-focal actor gender match on preference for third party intervention of the following nature: a. Female respondents will show *less* of a preference for third party intervention when paired with female non-focal actors as compared to when they are paired with male non-focal actors; b. Male respondents will show *less* of a preference for third party intervention when paired with female non-focal actors as compared to when they are paired with male non-focal actors.

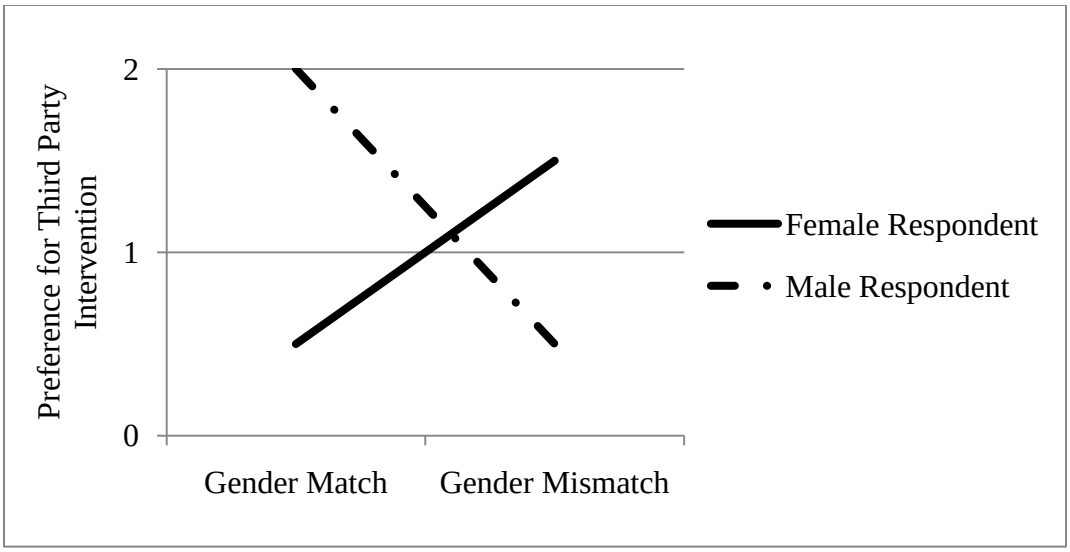

Figure 13. **Illustration of Hypothesis 12.**

Race *and* sex considerations in conflict. As discussed earlier, individuals belong to multiple groups (race, gender), and it is necessary to consider if individuals will use race, gender, or both to relate to actors in the conflict situational judgment items. It was also argued that context needs to be considered as it can influence the salience of various identities (again, Ridgeway & Smith-Lovin, 1999 provide a review). In the context of conflict situations, like in the context of helping situations, certain stereotypes may be used to make judgments about the non-focal actors in SJT items. When a conflict is based on a disagreement about the way work is getting done, one might consider the competence of the other party. The competence stereotype, it was argued earlier, has high race and high sex salience. When a conflict is based on interpersonal friction, it is less clear what stereotypes may be used to make inferences about the other party. However, it may be reasonable to expect that one might consider the agreeableness and aggressiveness of the other party. Both agreeableness-related and aggressiveness-related stereotypes arguably have high race and high sex salience. Given that Asians are seen as valuing relational harmony over personal interests (Brett & Kopelman, 2004, as cited by Liu, 2009), they

may be stereotyped as being more agreeable than other groups. Women are stereotyped as valuing relational harmony more so than men (Wilson et al., 2003) and therefore might be expected to be more agreeable than men. When it comes to aggressiveness, Blacks are stereotyped as being more aggressive than other groups (e.g., Harrison & Esqueda, 2001), and women are expected to be less aggressive than men (Huddy & Terkildsen, 1993).

Given that both types of conflict situations in the present study (disagreement about the way work is getting done, interpersonal friction) are likely to elicit stereotypes with high race and high sex salience, both race and sex of the target should be salient to the respondent. Predictions are therefore made with both expected race and expected sex effects in mind. First, considering assertiveness, the *lowest* level of assertiveness in responses is likely to be exhibited when females respond to non-focal actors of the same gender but of a different race, as compared to other sex and race-based combinations of respondents with targets. This expectation is based on the research already discussed. As proposed earlier, female respondents should respond to conflict with lower assertiveness than male respondents (Hypothesis 9a), and respondents are expected to be less assertive with non-focal actors of a different race (Hypothesis 6a). The *highest* level of assertiveness is likely to be exhibited by male respondents paired with non-focal actors of the same gender and race. This prediction is based on the expectations that men will be more assertive than women (Hypothesis 9a), that men will respond to male non-focal actors more assertively than to female non-focal actors (Hypothesis 10a), that more assertive responses will be seen in responses to those of the same race (Hypotheses 6a).

Hypothesis 13a: In conflict situations the lowest level of assertiveness will be exhibited by female respondents paired with a female non-focal actor of a different race as compared to other sex and race-based combinations of respondents with targets.

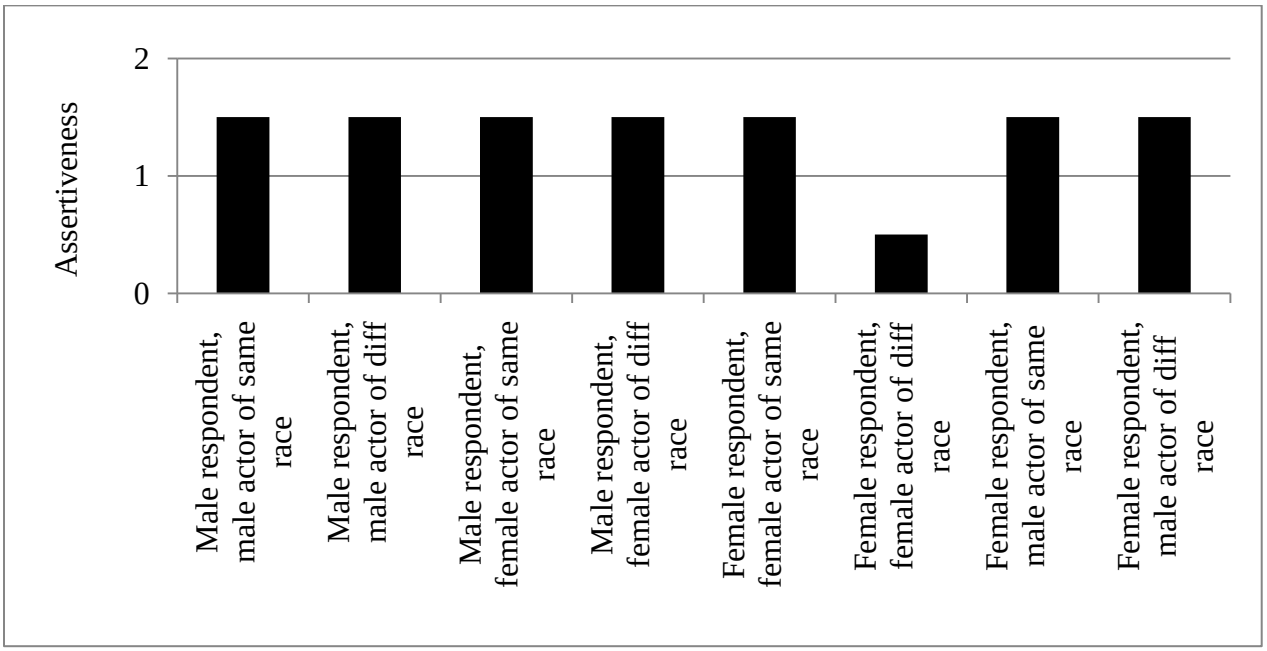

Figure 14. **Illustration of Hypothesis 13a.**

Hypothesis 13b: In conflict situations the highest level of assertiveness will be exhibited by male respondents paired with a male non-focal actor of the same race as compared to other sex and race-based combinations of respondents with targets.

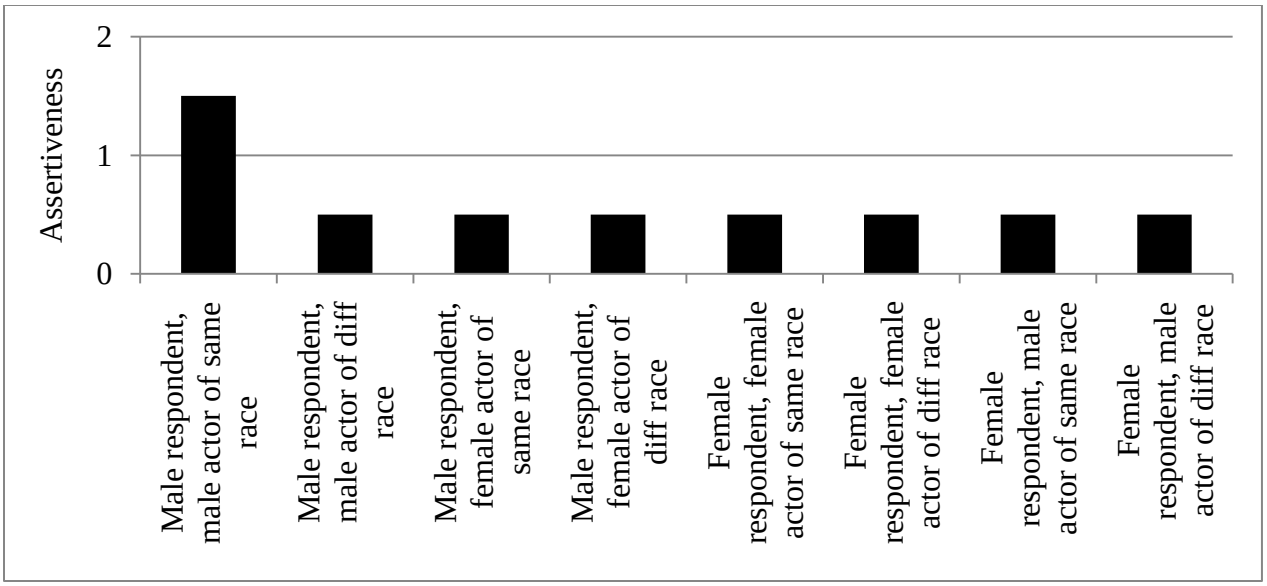

Figure 15. **Illustration of Hypothesis 13b.**

Next, considering cooperativeness, the *lowest* level of cooperativeness should be seen when male respondents are paired with non-focal actors of the same gender but a different race as compared to other sex and race-based combinations of respondents with targets. This expectation is based on research discussed earlier. It was proposed earlier that men will show lower cooperativeness than women (Hypothesis 9b), and that respondents will be less cooperative with those of a different race (Hypothesis 6b). The *highest* level of cooperativeness should be seen when female respondents are paired with non-focal actors of the same gender and race as compared to other sex and race-based respondent-target combinations. This expectation is also based on propositions put forth earlier. Specifically, women are expected to be more cooperative than men (Hypothesis 9b) and to respond to female non-focal actors with higher cooperativeness than to males (Hypothesis 10b). Respondents are expected to show more cooperativeness toward those of the same race as them (Hypothesis 6b).

Hypothesis 14a: In conflict situations the lowest level of cooperativeness will be exhibited by male respondents paired with male non-focal actors of a different race as compared to other sex and race-based combinations of respondents with targets.

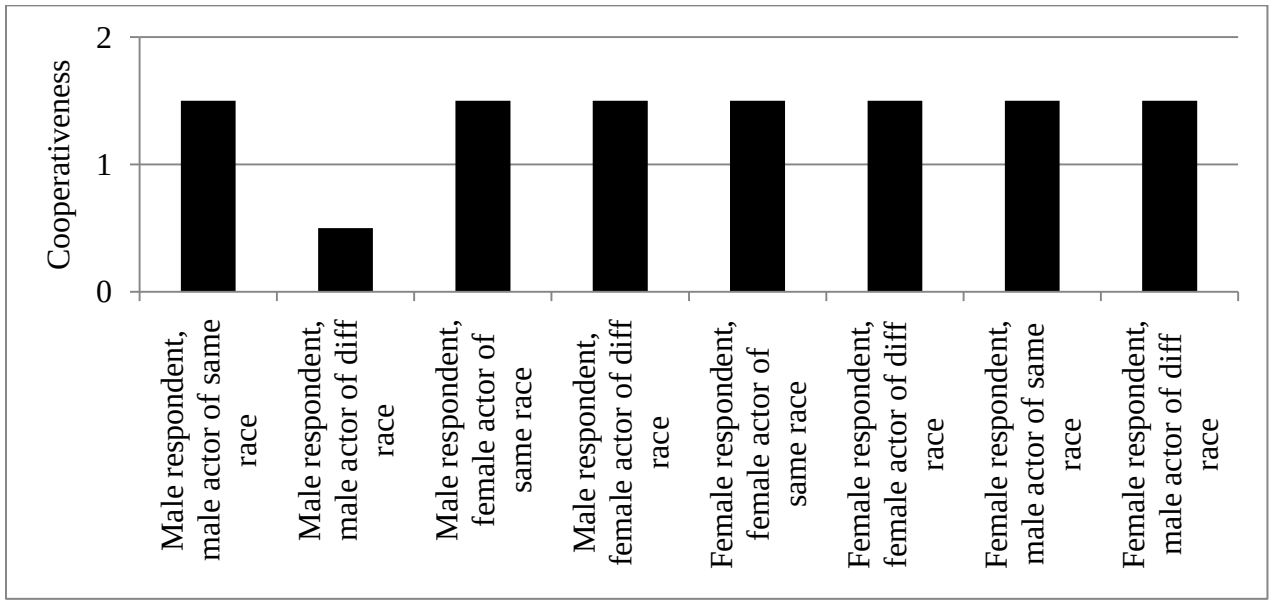

Figure 16. **Illustration of Hypothesis 14a.**

Hypothesis 14b: In conflict situations the highest level of cooperativeness will be exhibited by female respondents paired with a female non-focal actor of the same race as compared to other sex and race-based combinations of respondents with targets.

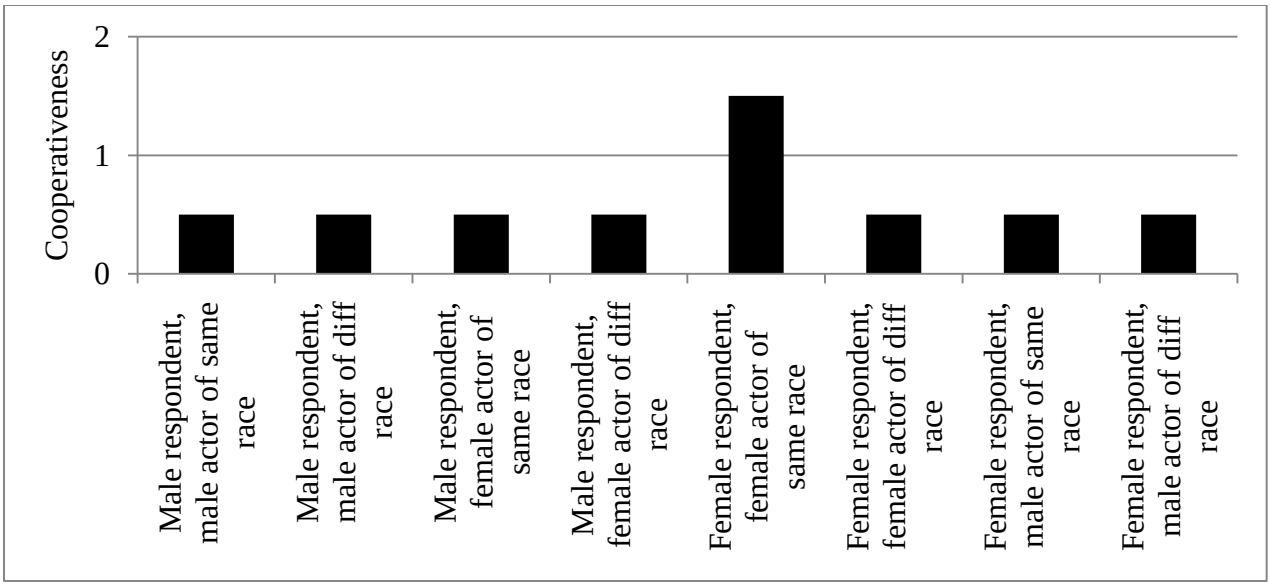

Figure 17. **Illustration of Hypothesis 14b.**

Finally, considering preference for third party intervention in the conflict situation, different expectations apply for the two types of situations considered (interpersonal friction, where the goal of third party involvement would be reconciliation; disagreement about the way work is getting done, where the goal of third party involvement would be reporting a coworker's poor performance). When the goal is reconciliation, the highest preference for third party intervention should be exhibited when female respondents from groups considered collectivist are paired with non-focal actors of the same gender and race. This hypothesis is based on the expectation that respondents from collectivist groups will show more of a preference for third party intervention when paired with non-focal actors of the same race (Hypothesis 7) and that respondents will show more preference for third party intervention when paired with female non-focal actors (Hypothesis 11).

Hypothesis 15: In a conflict situation where the goal of involving a third party is reconciliation, the highest preference for third party intervention will be exhibited by

collectivist (Asian, Black) respondents paired with a female non-focal actor of the same race.

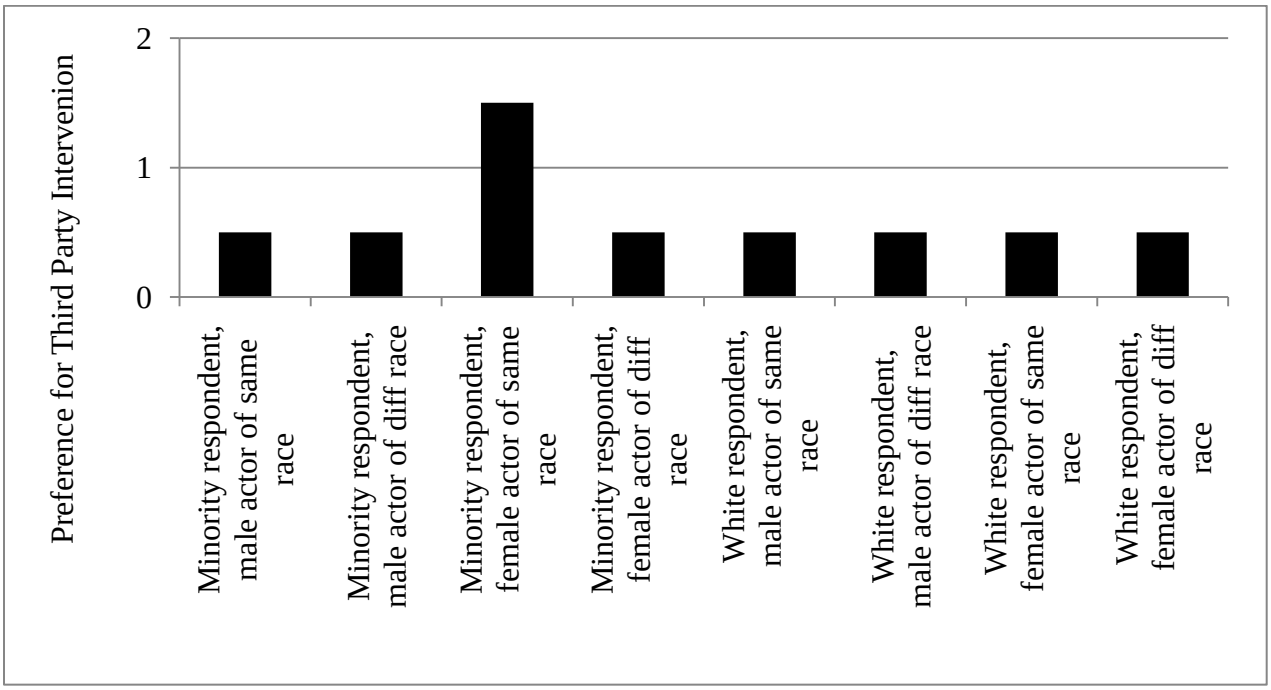

Figure 18. **Illustration of Hypothesis 15.**

When the goal is reporting a coworker's poor performance, the highest preference for third party intervention should be exhibited when male respondents from groups considered collectivist are paired with non-focal actors of a different race. This hypothesis is based on the expectation that respondents from collectivist groups will show more of a preference for third party intervention when paired with non-focal actors of a different race (Hypothesis 8) and that respondents will show more of a preference for an intervention when the non-focal actor is male (Hypothesis 12).

Hypothesis 16: In a conflict situation where the goal of involving a third party is reporting a coworker's poor performance, the highest preference for third party intervention will be exhibited by collectivist (Asian, Black) respondents paired with a male non-focal actor of a different race.

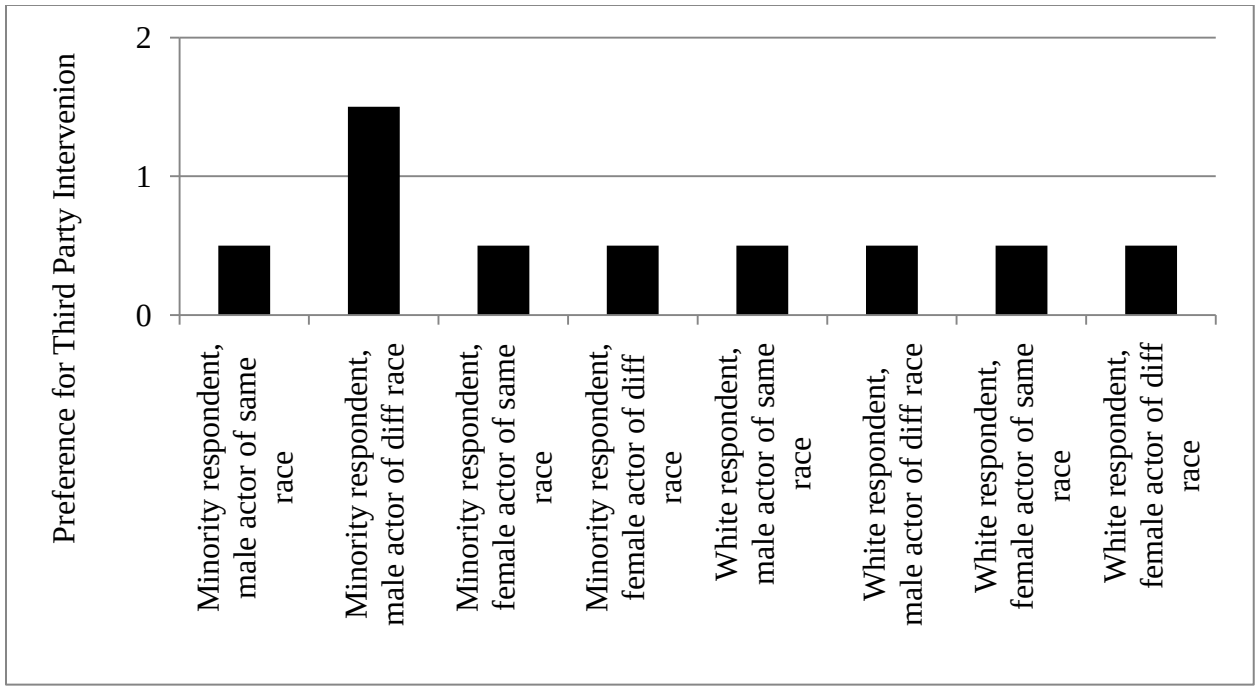

Figure 19. **Illustration of Hypothesis 16.**

Applicant Reactions

Applicant reactions to selection procedures are an important area of research and one with considerable implications for organizations. Organizational attractiveness, and intentions to accept a job offer, recommend the company to other people, undertake legal challenges, and buy company products have all been tied to applicant reactions (Hausknecht, Day, & Thomas, 2004). Meta-analytic findings show no significant main effects of race or sex on applicant reactions (e.g., perceptions of the fairness of the process and the outcomes received) (Hausknecht et al.,

2004). In the related research area on reactions to organizational recruitment there is evidence to show that the race and sex of individuals included in recruitment campaigns as representatives of the types of people in the hiring organization *relative to* the race and sex of the applicant does have an effect on applicant reactions. Researchers in the area of targeted recruiting have devoted substantial attention to examining the determinants of minority and female applicants' reactions to recruitment tactics (see Avery & McKay, 2006 for a review). Research has found that minorities (Blacks, Hispanics) react positively to racial diversity in advertisements while Whites are little influenced by level of diversity in ads (Avery, Hernandez, & Hebl, 2004; Perkins, Thomas, &Taylor, 2000). While less research has been done to evaluate Asians' reactions to diversity, they have been shown to react more positively toward ads featuring Asians that toward ads featuring Whites (Martin, Lee, & Yang, 2004). Notably, reactions toward diversity are likely to be influenced by attitudes viewers of ads hold (e.g., level of racial identification, other-group orientation, level of prejudice) (Avery, 2003; Whittler, 1991).

Interestingly, minorities (Blacks, Hispanics) seem to judge diversity based on minority group representation, as opposed to simply own group representation, reacting positively as long as either Blacks or Hispanics are well represented (Avery et al., 2004). These findings may arguably be extended to predict that these groups would also react positively to the presence of Asians, another minority group, and that Asians would react positively toward representation of these other groups in ads. White respondents' reactions, in the meantime, should not be affected by the race of the non-focal actors.

Hypothesis 17: Minority (Black, Asian) respondents will react more positively when the SJT features Black or Asian non-focal actors than when the SJT features White non-focal actors.

Like racial minorities are in relation to Whites, women are more attuned than men to diversity information (Avery & McKay, 2006). As explained later, the focal actor in the videos will always be female to try to simplify somewhat the design of the study. Given that the focal actor will always be female, in cases where the non-focal actor is also female, there will actually be no gender diversity (because of the absence of men). Therefore, women will probably react more positively to the SJT when the non-focal actors are male.

Hypothesis 18: Female respondents will react more positively when the non-focal actors are male.

In summary, existing research in the areas of relational demography, helping, and conflict suggests that different groups of test takers may respond to video-based situational judgment items differently depending on the race and gender of the actors in the videos. Research on applicant reactions to diversity in organizational recruitment materials suggests that test taker reactions to the overall test may likewise differ as a function of actor demographics. Next, I describe the method that was used to test the helping-, conflict-, and reactions-related hypotheses reviewed above.

METHOD

Design

The video-based SJT included two types of scenarios (help seeking, conflict) with two scenarios per scenario type. Thus, participants saw four scenarios, the order of which was counterbalanced.

The demographic characteristics of the focal actor (whose role the respondent was asked to assume) in each scenario were kept constant across study conditions. The focal actor was always a young White female. This was done in the interests of keeping the study design manageable. Though there is no guidance on this in the literature of which I am aware, the expectation is that the race and gender of the focal actor had little influence on the ability of the respondent to assume that individual's role when responding to SJT items. That is, there is little reason to believe that a respondent would actually try to "get inside the head" of the person whose role he or she needed to assume when determining what course of action to take as opposed to simply putting him or herself in the focal actor's shoes. It is more reasonable to expect that the respondent approached the situation from his or her own perspective when determining the best response as that is what respondents were explicitly instructed to do. If the respondents did in fact approach each situation from their own perspective, the race and gender of the focal actor should have mattered little. However, each focal actor was White and female, given that most study participants were expected to be White females (because of the composition of the psychology participant pool).

Across study conditions, the White female playing the focal person in a given SJT item was always the same actress. The non-focal actor, however, was varied to create six different versions of each of the four scenarios. That is, any given respondent saw four scenarios in all of which the non-focal actor was of one race and gender. Yet, across the four scenarios that this

respondent saw, the non-focal actors were played by different actors. To illustrate, a respondent might have seen 4 scenarios in which he/she always assumed the role of a White female to respond to a coworker who was always an Asian male, but across the 4 videos, all the actors were different Asian males.

The study assumed a 3x2x2x2 design (race of respondent: White, Black, Asian; race match between respondent and non-focal actor: yes, no; sex of respondent: male, female; sex match between respondent and non-focal actor: yes, no). The age of the non-focal person (as well as the focal person's, actually) was kept approximately constant, also in the interests of making the design somewhat more manageable.

Sample

The sample for this study was drawn from the psychology participant pool at Michigan State. The sample consisted of 374 students, of which 51% were female. The racial breakdown was as follows: 72.5% White, 11% Asian, 9.4% African American/Black, .3% Native Hawaiian or other Pacific Islander, 4.3% multi-racial, and 2.7% "other". Average age was 19.97 (SD = 2.68).

Procedure

Students came into a computer lab to participate in the study. A research assistant briefly instructed them on what they would be doing and handed out consent forms for them to sign. After reading and signing their consent forms students proceed to take the four-item SJT test. The test was preceded with several paragraphs asking students to imagine themselves as job applicants taking an interpersonal test as an initial step in the process of applying for a technician position at an attractive organization. This description was adapted from Bauer and colleagues (2001) who used it as a manipulation of organizational attractiveness with a college student

sample, and modified to fit the present study. The version used in this study is provided in Appendix A. Along with this manipulation respondents were also presented with a job description for an entry level plant technician.

Participants viewed the four SJT items on a computer screen, in medialab. Each video was followed by a screen where students indicated their likelihood of responding to the situation just viewed in each of five or six different ways, and what they were most and least likely to do. After completing the SJT students answered questions about their reactions. The order of the reactions measures (opportunity to perform, perceived job relatedness, attraction, compatibility) was counterbalanced such that half of the participants completed the opportunity to perform and perceived job relatedness measures first and the other half completed the attraction and compatibility measures first. This was done in case considering the fairness of the test should influence evaluations of the hypothetical organization or vice versa. Following the reactions measures, participants answered questions about their demographic backgrounds, rated the attractiveness of the actors, and indicated these actors' race and gender. Upon completion, students received a sheet with information about situational judgment tests (the required educational portion of the study).

Measures

Situational Judgment Test (SJT) Scenarios. Four scenarios were chosen from a full video-based interpersonal abilities SJT used in the past by a large, global corporation as part of its selection process for plant technicians. The test was developed to represent typical work situations encountered by plant technicians. Incumbent technicians provided critical incidents along with effective and ineffective responses to each. A concurrent validation design was used to establish the effectiveness of the test for predicting plant technician performance. The test was

validated against performance ratings for approximately 400 technicians. Ratings were provided by technicians' immediate supervisors.

All four SJT scenarios selected for this study dealt with interpersonal situations occurring between coworkers in an organization. More specifically, two scenarios pertained to help seeking and two pertained to interpersonal conflict. The scripts for each of these four scenarios and the accompanying questions are presented in Appendix C. Notes are included identifying which stereotypes are pertinent to each scenario as well as what the goal of intervention would be for those situations that pertain to conflict.

The scripts of these four scenarios were used to film six different versions of each of these scenarios. The twenty four scenario clips were filmed with the assistance of WKAR, Michigan State's on-campus TV station. Graduate and undergraduate students were recruited to act in these clips and were rewarded with $20 gift cards ($40 for focal actors). Once the clips were filmed, medialab was used to organize them into a form that was shown to study participants. Each scenario was followed by a screen where the respondent was asked to indicate the likelihood of his/her responding to the situation in each of five or six different ways. Likelihood ratings were provided for every option on a scale anchored by (1) *not at all likely* and (5) *very likely*. Respondents also chose one option they would *most likely* do and one option they would *least likely* do (e.g., "If you were Jane, what would you MOST likely do?" and "If you were Jane, what would you LEAST likely do?")

SJT scoring. There are three approaches to scoring and examining responses to the SJT in this study. The first, and the one most suited for testing the hypotheses in this study, is by weighting likelihood ratings test takers gave to each response option according to the level of helping, assertiveness, cooperation, or third party intervention a given response option

represents. Weighted likelihood ratings for the set of response options can then be summed up to form scores representing helping, assertiveness, cooperativeness, and third party intervention. Each test taker would then have scores on each of these dimensions: 2 helping scores for the two helping SJT items and 2 assertiveness, 2 cooperativeness, and 2 third party intervention scores for the two conflict SJT items.

To determine the level of relevant construct(s) represented by each SJT response option (i.e., the item weights), judges were asked to provide ratings of options. They were recruited on an individual basis through personal appeals and were unaware of the study's hypotheses. The final sample consisted of 40 judges. They were graduate students (51.3%), college graduates (25.6%), college students (20.5%), and doctorate degree holders (2.6%). One judge did not indicate level of education. They were almost evenly split in terms of gender (52.5% male). Judges were White (72.5%), Asian (15%), Black (2.5%), and other/multi-racial (10%).

These individuals rated response options to the two conflict items on level of assertiveness [using a scale from (1) *very unassertive* to (5) *very assertive*], cooperativeness [using a scale from (1) *very uncooperative* to (5) *very cooperative*], and third party intervention [using a scale from (1) *none* to (5) *very high amount*]. They rated response options to the two helping items on level of helping [using a scale from (1) *very unhelpful* to (5) *very helpful*].

A second way to score, or in this case simply *examine*, responses to the SJT is to, again, use the likelihood ratings test takers assigned to response options, but not weight these by the corresponding level of a construct of interest (e.g., assertiveness). When this approach is used, any analyses of responses are done at the level of the individual response options. If differences between [groups of] test takers emerge, one has to take a close look at the individual response options to come up with a substantive explanation of why likelihood ratings on these response

options should differ between [groups of] test takers. This is a different type of approach to explanation than scoring method 1 (which allows one to look across response options in analyses and speak about the *dimensions* on which responses of different [groups of] test takers may differ). As such, method 2 is somewhat less amenable to testing the dimension-based hypotheses in this study.

A third way to score the SJT is according to response effectiveness as determined by the test developer. Respondents' most and least likely responses to the four SJT items were scored according to the scoring program used by the organization that provided their SJT for this study. A *most likely* (ML) response was assigned a score of 1 if it matched what the company considered to be a desirable response to the situation, a score of -1 if it matched a response considered an undesirable option, and a score of 0 if it was a match for neither a desirable nor an undesirable response. A *least likely* (LL) response was assigned a score of 1 if it matched what the company considered to be an undesirable response to the situation, a score of -1 if it matched a desirable response, and a score of 0 if it was a match for neither an undesirable nor a desirable response. To clarify what defined "desirable" and "undesirable" it should be noted that the test developer based the scoring key on response frequencies obtained from the sample of technicians used to validate the test. Responses chosen most often by incumbents (specifically, by 100 or more respondents) were considered desirable responses and received the highest scores. Responses chosen least often by incumbents (specifically, by 0-25 respondents) were considered undesirable responses and received the lowest scores. The zero scores were assigned to response options chosen by 26 to 99 technicians in the sample. The resultant scoring key was reviewed by personnel at the company to make sure that the scoring of all the response options was consistent with performance expectations for plant technicians.

In this study the scores for the most likely response options were added up across the 4 SJT items to come up with an ML score, and scores for the least likely response options were added up across the 4 SJT items to come up with an LL score. The ML and LL scores were then combined to get a total SJT test score for each study participant. This approach to scoring the SJT is most appropriate for examining *performance* on the SJT and identifying any differences between groups of test takers that may exist. Similar to method 2 described above, the most and least likely responses test takers provided can be examined for differences in which response options were chosen most frequently by different groups. Thus, frequencies would be the dependent variables in analyses (versus dimensions in method 1 and likelihood ratings in method 2). Like with method 2 one would then look closely at the individual response options to suggest a substantive explanation for any differences. Another possibility is to separately use test takers' scores on the 4 scenarios as dependent variables; helping item scores could be used in examining hypotheses pertaining to helping and conflict item scores could be used in examining hypotheses pertaining to assertiveness, cooperativeness, and intervention.

All three methods of scoring and examining SJTs were applied to various extents when analyzing the results of this study.

Applicant reactions. Reactions to the SJT were operationalized in terms of perceptions of opportunity to perform, perceptions of job relatedness, and attraction to and compatibility with an organization administering such an SJT. Perceptions of opportunity to perform were measured with four items adapted from Bauer and her colleagues (2001). Answers were provided on a five point scale ranging from (1) *strongly disagree* to (5) *strongly agree*. A sample item is "This test allowed me to show what my interpersonal skills are." Reliability was $\alpha = .84$ for the measure. The items are provided in Appendix D.

Perceived job relatedness was measured with four items adapted from Bauer and colleagues (2001). Answers were provided on a five point scale ranging from (1) *strongly disagree* to (5) *strongly agree*. A sample item is "Doing well on this test means a person can do the technician job well." Bauer and colleagues found two of their job relatedness items loaded on one factor ("job relatedness—predictive") and two loaded on a second factor ("job relatedness—content"). Reliability was $\alpha = .81$ for the predictive items and $\alpha = .83$ for the content items. The items are provided in Appendix E.

Respondents were asked to indicate their attraction to the organization to which they were hypothetically applying. Attraction was assessed with six items adapted from Perkins and colleagues (2000). A sample item is "I would speak to a company representative about the possibility of employment." Individuals indicated their agreement with each item using a scale ranging from (1) *strongly disagree* to (5) *strongly agrees*. Reliability was $\alpha = .88$ for this scale. The items are provided in Appendix F.

Respondents were also asked to indicate their perceived compatibility with the hypothetical organization. Compatibility was assessed with four items adapted from Perkins and colleagues (2000) including "I would feel at home working for an organization like this." Individuals indicated their agreement with each item using a scale ranging from (1) *strongly disagree* to (5) *strongly agrees*. Reliability was $\alpha = .85$ for this scale. The items are provided in Appendix G.

Demographic questions. Respondents answered questions about their age, gender, ethnicity, race, year in college, GPA, and intended major. The list of demographic questions is provided in Appendix H.

Coworker attractiveness. Because physically attractive individuals are initially liked more than less attractive individuals (Berscheid & Walster, 1974), the attractiveness of coworkers (i.e., non-focal actors) was measured so that it could be controlled for in the analyses. Respondents were presented with photos of all the actors that appeared in the videos they watched and asked to rate the attractiveness of each on a scale of (1) *very unattractive* to (5) *very attractive*.

Manipulation checks. To make sure that the gender and race manipulations worked, respondents were presented with photos of all the actors that appeared in the videos they watched and asked to indicate the race and gender of each. The response options were Asian female, Asian male, African-American/Black female, African-American/Black male, Middle Eastern female, Middle Eastern male, White female, White male, and cannot tell. Even though none of the actors were Middle Eastern, this was included as a choice to make these questions seem more difficult (with the goal of discouraging respondents from picking unreasonable options just to spite the experimenter for asking for very obvious responses).

Familiarity with actors. At the end of the experiment, respondents were asked whether they were acquainted with any of the actors they saw in the four videos they watched (yes, no). Those who responded "yes" were presented with numbered photos of the individuals they saw in the videos and asked to indicate the number(s) of the individual(s) with whom they were acquainted.

Next, the results of the study are reviewed. Tests of the hypotheses presented earlier as well as exploratory analyses are described.

RESULTS

Manipulation Checks

Twenty nine respondents of the 374 failed to identify the race and/or gender of the non-focal actor in conflict situation 1, 28 of conflict situation 2, 12 of helping situation 1, and 27 of helping situation 2. It should be noted that I did not consider identifying a Caucasian non-focal actor as Middle Eastern to be a failure of the manipulation check because Middle Easterners are technically Caucasian. Individuals who failed the manipulation check for a particular situation were not included in the analysis of that situation (e.g., respondents who failed the manipulation check for help seeking situation 2 were not included in analyses of help seeking situation 2).

Analyses were done to try to understand why so many respondents failed one or more of the manipulation checks. As a group, those who failed at least one manipulation check were not different from those who passed all the manipulation checks in terms of race [$\chi^2(5, N = 374) = 2.37$, ns], gender [$\chi^2(1, N = 374) = 1.39$, ns], or age [$t(371) = .08$, ns]. Though there were no demographic differences between the groups, there were differences in response latencies on the manipulation checks. Respondents who failed the manipulation check for conflict situation 1 spent significantly more time (recorded in milliseconds) on the question asking them to indicate the demographic profile of the non-focal actor (i.e., the manipulation check) ($M = 9993.24$ ms, $SD = 5852.47$) than those who passed the manipulation check for that situation ($M = 5608.03$ ms, $SD = 4031.25$), $t(30.27) = 3.96$, $p < .001$. Respondents who failed the manipulation check for conflict situation 2 spent significantly more time on the question ($M = 6008.18$ ms, $SD = 2847.97$) than those who passed the manipulation check for that situation ($M = 2948.95$ ms, $SD = 2093.55$), $t(372) = 7.22$, $p < .001$. Respondents who failed the manipulation check for helping situation 1 spent marginally more time on the question ($M = 4567.92$ ms, $SD = 2629.83$) than

those who passed the manipulation check for that situation (M = 2894.47 ms, SD = 1953.75), $t(11.41) = 2.18$, $p = .051$. Finally, respondents who failed the manipulation check for helping situation 2 spent significantly more time on the question (M = 4455.07 ms, SD = 2434.11) than those who passed the manipulation check for that situation (M = 2219.59 ms, SD = 1327.90), $t(27.22) = 4.72$, $p < .001$. This pattern of differences suggests that respondents who failed the manipulation check(s) were not unduly quick in identifying the actors (as opposed to tending to simply rush through the questions and making errors because of that).

The actual responses of those who failed the manipulation check(s) were examined to determine the nature of the identification difficulties their longer response times suggest. Table 1 provides a summary of the nature of respondents' demographic profile identification difficulties across the four manipulation checks. Overall, respondents had the most trouble correctly identifying the demographic profile of the Black male and the White non-focal actors. Across conditions the Middle Eastern response option was the biggest distracter for those who did not correctly identify the demographic profiles of non-focal actors. The fact that some respondents seemed to have misidentified actors' gender (due to inattentiveness most likely, as opposed to genuine identification difficulties, as none of the actors' sex was ambiguous) also contributed to manipulation check failures. Tables in Appendix I show the breakdown of incorrect responses to the manipulation checks for the 4 SJT items individually. As a reminder, individuals who failed the manipulation check for a particular situation were not included in the analysis of that situation.

Table 1. **Distribution of incorrect responses across manipulation checks**

Condition	East Asian Female	East Asian Male	Black Female	Black Male	Caucasian Female	Caucasian Male	Middle Eastern Female	Middle Eastern Male	Cannot tell	Total *N*
Asian Female non-focal	X	7.1%	--	--	--	--	64.3%	--	28.6%	14
Asian Male non-focal	**16.7%**	X	--	--	--	--	--	50%	33.3%	12
Black Female non-focal	--	--	X	**33.3%**	--	--	33.3%	--	33.3%	3
Black Male non-focal	--	4.3%	**4.3%**	X	--	--	--	56.5%	34.8%	23
White Female non-focal	8.3%	--	--	--	X	**4.2%**	--	--	87.5%	24
White Male non-focal	--	5%	--	--	**30%**	X	--	--	65%	20

Note. Bolded entries indicate errors based solely on gender error.

Familiarity with Actors

Thirty four respondents indicated knowing at least one of the actors in the videos. Specifically, for conflict situation 1, 7 knew the focal actor and 10 knew the non-focal actor. For conflict situation 2, 9 knew the focal actor and 5 knew the non-focal actor. For helping situation 1, 7 knew the focal actor and 3 knew the non-focal actor. For helping situation 2, 2 knew the focal actor and 13 knew the non-focal actor. As discussed later, familiarity with actors did not have much influence on test takers' responses.

Item Weights

In order to aggregate judges' perceptions of what level of a particular construct (e.g., cooperativeness, assertiveness, third party intervention, helping) each SJT response option represents, and using means to represent collective perceptions of response options, an adequate level of inter-rater agreement (indicator of the extent to which individuals provided the same ratings) had to be demonstrated (Kozlowski & Hattrup, 1992). James et al.'s (1984) r_{WG} index for assessing agreement among a set of judges rating one target using a single item was used to assess agreement for each of the 44 items (ratings of one response option along one dimension represent a single item). Before estimating r_{WG} values, ratings that were more than three standard deviations away from the mean of any given item were judged to be outliers and removed. Sixteen outliers were removed from the data. A uniform null distribution was used to estimate r_{WG} values. Due to initially low estimates of r_{WG}, the decision was made to collapse the 5 point scale used by judges into a three point scale. For cooperativeness, assertiveness, and helping ratings, ratings of 1 and 2 were collapsed into 1's (low), ratings of 3 were recoded into 2's (neutral), and ratings of 4 and 5 were collapsed into 3's (high). For third party intervention

ratings, ratings of 1 remained as 1's (none), ratings of 2 and 3 were collapsed into 2's (medium) and ratings of 4 and 5 were collapsed into 3's (high). Using the collapsed 3-point scales to estimate r_{WG} resulted in values ranging from .53 to 1.00 with a mean of .79. These r_{WG} values along with the means and standard deviations of judges' ratings are presented in Table 2 (conflict response options) and Table 3 (helping response options). Using the recommendations provided by LeBreton and Senter (2008) to interpret values of .51 to .70 as moderate agreement, .71 to .90 as strong agreement, and .91 to 1.00 as very strong agreement, 27.3%, 45.5%, and 27.3% the r_{WG} values fall within these categories, respectively (note that these percentages are computed based on the 44 r_{WG} values in Tables 2 and 3 combined). Thus, there was adequate agreement among judges to use the means of ratings they provided.

Test takers' likelihood ratings of SJT response options were consequently weighted by the appropriate mean ratings provided by the sample of judges to come up with an assertiveness, cooperativeness, and intervention score for each test taker for each of the response options associated with the two conflict SJT items, and a helping score for each of the response options associated with the two helping SJT items. Response option scores for each item were added up to come up with an item-level score for each person (2 cooperativeness scores, 2 assertiveness scores, and 2 intervention scores corresponding to the two conflict SJT items; 2 helping scores corresponding to the two helping SJT items). Item-level scores were also summed up to come up with test-level scores for test takers (i.e., overall scores for cooperativeness, assertiveness, intervention, and helping).

Table 2. **Means, standard deviations, and r_{WG} values for judges' ratings of conflict response options**

Response Option	Dimension								
	Cooperativeness			Assertiveness			Intervention		
	M	*SD*	r_{WG}	*M*	*SD*	r_{WG}	*M*	*SD*	r_{WG}
CS1 Option 1: Tell the coworker that you will continue to watch for unglued cartons and get back to him/her.	2.55	.78	.69	1.88	.91	.58	1.08	.27	.96
CS1 Option 2: Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.	2.55	.64	.80	1.10	.31	.95	1.11	.31	.95
CS1 Option 3: Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.	1.45	.75	.72	2.95	.32	.95	1.27	.61	.82
CS1 Option 4: Ask another team member to discuss the problem with your coworker.	2.13	.76	.71	2.35	.74	.73	2.80	.41	.92
CS1 Option 5: Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.	1.30	.65	.79	2.85	.49	.88	2.70	.61	.82
CS2 Option 1: Ask your coworker what you could do to get along better.	2.92	.35	.94	2.08	.97	.53	1.11	.40	.92
CS2 Option 2: Ask the team leader for suggestions about how to work with your coworker.	2.60	.67	.77	2.15	.89	.60	2.77	.54	.86
CS2 Option 3: Try to avoid your coworker whenever possible because you can't change him/her.	1.55	.68	.77	1.67	.84	.65	1.00	.00	1.00
CS2 Option 4: Tell your coworker that he/she should act like a team member and try to work together.	1.73	.93	.56	2.75	.63	.80	1.22	.48	.89
CS2 Option 5: Ask the team leader to meet with you and your coworker to work out your differences.	2.43	.75	.72	2.53	.78	.69	2.95	.22	.98
CS2 Option 6: Ask other team members for suggestions on how to work with your coworker.	2.40	.84	.65	2.33	.83	.66	2.81	.40	.92

Note. CS1 = Conflict situation 1. CS2 = Conflict situation 2.

Table 3. **Means, standard deviations, and r_{WG} values for judges' ratings of helping response options**

Response Option	Helping Dimension		
	M	*SD*	r_{WG}
HS1 Option 1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.	2.30	.85	.64
HS1 Option 2: Suggest that your coworker ignore Frank because working with Frank can be difficult.	1.43	.75	.72
HS1 Option 3: Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank.	1.30	.61	.82
HS1 Option 4: Tell your coworker how you deal with Frank.	2.97	.16	.99
HS1 Option 5: Suggest that your coworker meet with Frank to talk about their differences.	2.80	.52	.87
HS1 Option 6: Suggest that your coworker talk to the team leader about his/her problem with Frank.	2.50	.72	.74
HS2 Option 1: Agree to switch with your coworker so that he/she can take the training course.	2.70	.65	.79
HS2 Option 2: Agree to switch with your coworker, but only if he/she will switch with you in a few weeks.	2.95	.22	.98
HS2 Option 3: Tell your coworker you think he/she is taking advantage of the system and you won't switch.	1.33	.66	.78
HS2 Option 4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	1.63	.81	.67
HS2 Option 5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	2.33	.86	.63

Note. HS1 = Helping situation 1. HS2 = Helping situation.

As an alternative to collapsing the 5-point response scale into a 3-point scale to improve agreement between judges' ratings of the response options, I also tried to standardize ratings within judge for each SJT item with rated dimension combination (e.g., for helping ratings given to response options of helping item 1) before computing mean ratings. This procedure resulted in adequate r_{WG} values as seen in Tables 4 and 5. However, I proceeded to use the procedure described earlier as it was more convenient to interpret means on a 3-point scale (as opposed to means of Z-scores) in the context of substantive analyses of test takers' likelihood ratings of these response options. Additionally, using Z-score weights did not improve the reliability issue described later. It should be noted that using this approach would eliminate support for Hypothesis 3 (see "H3, scoring method 1").

Table 4. **Means, standard deviations, and r_{WG} values for judges' standardized ratings of conflict response options**

Response Option	Dimension								
	Cooperativeness			Assertiveness			Intervention		
	M	*SD*	r_{WG}	*M*	*SD*	r_{WG}	*M*	*SD*	r_{WG}
CS1 Option 1: Tell the coworker that you will continue to watch for unglued cartons and get back to him/her.	.54	.66	.78	-.29	.62	.81	-.71	.25	.97
CS1 Option 2: Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.	.67	.87	.62	-1.28	.41	.92	-.69	.23	.97
CS1 Option 3: Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.	-.56	.70	.76	.89	.34	.94	-.54	.40	.92
CS1 Option 4: Ask another team member to discuss the problem with your coworker.	.11	.61	.81	.07	.50	.88	1.02	.39	.92
CS1 Option 5: Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.	-.76	.58	.83	.63	.44	.90	.83	.72	.74
CS2 Option 1: Ask your coworker what you could do to get along better.	.99	.51	.87	-.22	.79	.69	-.85	.30	.96
CS2 Option 2: Ask the team leader for suggestions about how to work with your coworker.	.31	.47	.89	-.08	.73	.73	.66	.44	.90
CS2 Option 3: Try to avoid your coworker whenever possible because you can't change him/her.	-.93	.79	.69	-.83	1.08	.42	-.96	.22	.98
CS2 Option 4: Tell your coworker that he/she should act like a team member and try to work together.	-.54	.91	.59	.68	.80	.68	-.74	.41	.92
CS2 Option 5: Ask the team leader to meet with you and your coworker to work out your differences.	.08	.67	.78	.29	.72	.74	1.01	.30	.96
CS2 Option 6: Ask other team members for suggestions on how to work with your coworker.	.11	.66	.78	.13	.57	.84	.77	.36	.94

Note. CS1 = Conflict situation 1. CS2 = Conflict situation 2.

Table 5. **Means, standard deviations, and r_{WG} values for judges' standardized ratings of helping response options**

Response Option	Helping Dimension		
	M	*SD*	r_{WG}
HS1 Option 1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.	.11	.75	.72
HS1 Option 2: Suggest that your coworker ignore Frank because working with Frank can be difficult.	-.82	.71	.75
HS1 Option 3: Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank.	-1.01	.63	.80
HS1 Option 4: Tell your coworker how you deal with Frank.	.76	.38	.93
HS1 Option 5: Suggest that your coworker meet with Frank to talk about their differences.	.71	.55	.85
HS1 Option 6: Suggest that your coworker talk to the team leader about his/her problem with Frank.	.29	.50	.88
HS2 Option 1: Agree to switch with your coworker so that he/she can take the training course.	.54	.83	.66
HS2 Option 2: Agree to switch with your coworker, but only if he/she will switch with you in a few weeks.	.76	.36	.94
HS2 Option 3: Tell your coworker you think he/she is taking advantage of the system and you won't switch.	-.89	.64	.80
HS2 Option 4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	-.50	.68	.77
HS2 Option 5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	.09	.67	.78

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

As an alternative to using means for weighting likelihood ratings assigned to response options before computing composite scores, an attempt was made to instead apply unit weights of 1 and 2 for low and high levels of a dimension of interest. Determinations of low or high were made based on judges' mean rating for a given response option falling either below or above the midpoint on the collapsed 3-point rating scale. The weights assigned using this procedure are presented in Tables 6 and 7 below. Unit weights like Z-score weights did not improve the reliability issue described later and were ultimately not used. It should be noted, however, that using this approach would produce the same results for Hypotheses 1 through 5.

Table 6. **Unit weights associated with judges' standardized ratings of conflict response options**

Response Option	Weight for Cooperativeness	Weight for Assertiveness	Weight for Intervention
CS1 Option 1: Tell the coworker that you will continue to watch for unglued cartons and get back to him/her.	2.00	1.00	1.00
CS1 Option 2: Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.	2.00	1.00	1.00
CS1 Option 3: Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.	1.00	2.00	1.00
CS1 Option 4: Ask another team member to discuss the problem with your coworker.	2.00	2.00	2.00
CS1 Option 5: Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.	1.00	2.00	2.00
CS2 Option 1: Ask your coworker what you could do to get along better.	2.00	2.00	1.00
CS2 Option 2: Ask the team leader for suggestions about how to work with your coworker.	2.00	2.00	2.00
CS2 Option 3: Try to avoid your coworker whenever possible because you can't change him/her.	1.00	1.00	1.00
CS2 Option 4: Tell your coworker that he/she should act like a team member and try to work together.	1.00	2.00	1.00
CS2 Option 5: Ask the team leader to meet with you and your coworker to work out your differences.	2.00	2.00	2.00
CS2 Option 6: Ask other team members for suggestions on how to work with your coworker.	2.00	2.00	2.00

Note. CS1 = Conflict situation 1. CS2 = Conflict situation 2.

Table 7. **Unit weights associated with judges' standardized ratings of helping response options**

Response Option	Weight for Helping
HS1 Option 1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.	2.00
HS1 Option 2: Suggest that your coworker ignore Frank because working with Frank can be difficult.	1.00
HS1 Option 3: Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank.	1.00
HS1 Option 4: Tell your coworker how you deal with Frank.	2.00
HS1 Option 5: Suggest that your coworker meet with Frank to talk about their differences.	2.00
HS1 Option 6: Suggest that your coworker talk to the team leader about his/her problem with Frank.	2.00
HS2 Option 1: Agree to switch with your coworker so that he/she can take the training course.	2.00
HS2 Option 2: Agree to switch with your coworker, but only if he/she will switch with you in a few weeks.	2.00
HS2 Option 3: Tell your coworker you think he/she is taking advantage of the system and you won't switch.	1.00
HS2 Option 4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	1.00
HS2 Option 5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	2.00

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

Reliability of Measurement of Hypothesized SJT Item Dimensions

Reliability analysis performed on the weighted response options to verify that the expected dimensions of helping, assertiveness, cooperativeness, and third party intervention were being measured reliably by the appropriate weighted response options, showed internal consistency reliability to be low in all cases. For the overall dimensions, alphas ranged from .404 for the cooperativeness dimension to .518 for the helping dimension. When reliability was computed for the response options for each SJT item separately, reliability values ranged from .257 for intervention based on one item to .604 for helping based on one item. These reliabilities are shown in Table 8. Given that none of the alphas reached even the .70 minimum convention of acceptability, there was inadequate evidence for the reliable measurement of the dimensions I judged to be present in the response options. Given that the a priori dimensional structure was not producing reliable measures, the likelihood ratings for the entire set of response options (i.e., 22 options) was subjected to an exploratory factor analysis (EFA) to determine if a more interpretable structure existed. Maximum likelihood estimation with varimax rotation was used for the EFA. Eight factors were extracted and explained 38.38% of the variance. This model adequately fit the data, $\chi^2(84) = 90.04$, *ns*. Loadings of items on the factors are presented in Table 9. It should be noted that a forced solution of 4 factors (because there were four SJT items) and 2 factors (because there were items of 2 types) were also tried but the goodness of fit test [$\chi^2(132) = 223.34$, $p < .001$, and $\chi^2(169) = 398.15$, $p < .001$, respectively] indicated that these models did not fit the data.

Three of the response options associated with helping situation 2 showed high loadings onto the first factor, which explained 9% of the variance in the data. A close look at these three response options confirmed that the factor could be considered “helping”. Likelihood ratings for

option 1, option 3 (reverse coded such that a higher likelihood rating corresponded to a higher level of helping), and option 4 (also reverse coded) were therefore weighted by their associated standing on the helping dimension (as rated by 40 judges) and added up to form a helping score. Internal consistency reliability for this helping composite was .745. As can be seen in Table 9 an additional response option (5^{th} one) also loaded onto the first factor, however, it was not included in the helping composite as it would reduce reliability down to .67. The remaining items did not hang together strongly enough to warrant creating additional composite scores for subsequent analyses (alphas of no higher than .64).

Table 8. **Reliabilities of SJT item dimensions**

Dimension	Alpha
Cooperativeness, overall	.404
Cooperativeness for conflict item 1	.047
Cooperativeness for conflict item 2	.407
Assertiveness, overall	.464
Assertiveness for conflict item 1	.261
Assertiveness for conflict item 2	.382
Intervention, overall	.467
Intervention for conflict item 1	.257
Intervention for conflict item 2	.409
Helping, overall	.518
Helping for helping item 1	.394
Helping for helping item 2	.604

Table 9. **Loadings for likelihood ratings respondents gave to SJT response options**

Rated Response Option	Factor							
	1	2	3	4	5	6	7	8
HS2 O1: Agree to switch with your coworker so that he/she can take the training course.	-.881			-.207				.290
HS2 O3: Tell your coworker you think he/she is taking advantage of the system and you won't switch.	.708							
HS2 O4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	.683							
HS2 O5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	.321							.242
HS1 O6: Suggest that your coworker talk to the team leader about his/her problem with Frank.		.679						
CS2 O5: Ask the team leader to meet with you and your coworker to work out your differences.		.534						.214
CS2 O2: Ask the team leader for suggestions about how to work with your coworker.		.509				.475		
HS1 O5: Suggest that your coworker meet with Frank to talk about their differences.		.377			-.230			
HS1 O1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.		-.227						
CS1 O1: Tell your coworker that you will continue to watch for unglued cartons and get back to him/her.			.978					
HS2 O2: Agree to switch with your coworker, but only if he/she will switch with you in a few weeks.								

Table 9 (cont'd)

	1	2	3	4	5	6	7	8
CS1 O5: Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.				.595				
CS1 O4: Ask another team member to discuss the problem with your coworker.				.455				
HS1 O3: Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank.		-.210		.325				
CS1 O3: Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.				.294			.271	
HS1 O2: Suggest that your coworker ignore Frank because working with Frank can be difficult.					.542			
CS2 O3: Try to avoid your coworker whenever possible because you can't change him/her.					.530	-.213		
CS1 O2: Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.					.232			
CS2 O1: Ask your coworker what you could do to get along better.						.623		
HS1 O4: Tell your coworker how you deal with Frank.							.566	
CS2 O6: Ask other team members for suggestions on how to work with your coworker.							.408	
CS2 O4: Tell your coworker that he/she should act like a team member and try to work together.								.430
Variance explained	9.00%	6.14%	4.81%	4.57%	4.14%	3.73%	3.28%	2.72%

Note. $N = 374$. Factor loadings < .2 are suppressed. CS1 = Conflict situation 1. CS2 = Conflict situation 2. HS1 = Helping situation 1. HS2 = Helping situation 2.

Recall that there are a number of ways the SJT in this study can be scored (or analyzed). The first method (the one most suited for testing the hypotheses in this study), is by weighting likelihood ratings test takers gave to each response option according to the level of the construct of interest they represent and summing these weighted ratings into a dimension score (e.g., helping score). The dimension score can then be used as the dependent variable. The second method is to use the likelihood ratings test takers assigned to response options as is (i.e., not weight them), as separate dependent variables. The third method is to score the most and least likely responses test takers provided for each SJT item according to response effectiveness as determined by the test developer. The dependent variable is then the item score (or the test score if the four item scores are summed up).

Given the above discussion of reliability issues, scoring method 1 can only be applied for testing some of the helping hypotheses (i. e., the ones that include or pertain uniquely to helping situation 2 as none of the first helping situation's response options went into forming the "helping score"). The hypotheses were therefore *tested* using scoring method 1 whenever possible and *explored* (more so than tested) using the other two scoring methods when scoring method 1 could not be applied at all or could only be partially applied. It should be pointed out, however, that scoring method 3, as it is more appropriate for testing for effects on performance rather than on constructs like helping or assertiveness, is not at all suitable for testing my hypotheses pertaining to constructs. Method 3 is nonetheless applied and discussed in the interests of understanding demographic cue effects on SJT performance.

Correlations

The intercorrelations, means, standard deviations, and reliabilities for study variables are provided in Table 10 (note that explanation of variable names and other notes about the table are

presented on page 89). Inconsistent with the expectation that attractiveness of the non-focal actors will influence responses to the SJT items, ratings of attractiveness were not related to ratings of response options in expected ways. Thus, there was no strong support for controlling for non-focal actor attractiveness in the analyses. Correlations of variables representing knowing actors in the videos with responses on the SJT showed knowing actors to have very little influence (knowing the non-focal actor in conflict situation 1 was related to likelihood ratings for the first response option, r = -.12; knowing the non-focal actor in helping situation 1 was related to likelihood ratings for the third response option, r = .13). Thus, it did not seem necessary to control for knowing actors in the analyses.

Table 10. **Intercorrelations, means, standard deviations, and reliabilities**

	M	*SD*	1. CS1.1	2. CS1.2	3. CS1.3	4. CS1.4	5. CS1.5	6. CS2.1	7. CS2.2	8. CS2.3	9. CS2.4	10. CS2.5	11. CS2.6	12. HS1.1	13. HS1.2	14. HS1.3	15. HS1.4	16. HS1.5
1. CS1.1	4.29	.81	-															
2. CS1.2	2.51	1.05	-.03	-														
3. CS1.3	3.48	1.05	.01	-.07	-													
4. CS1.4	2.77	1.13	-.02	-.02	**.14**	-												
5. CS1.5	3.03	1.19	-.03	**-.17**	**.18**	**.28**	-											
6. CS2.1	3.92	1.10	.09	-.04	-.01	**.13**	.08	-										
7. CS2.2	3.72	1.08	**.11**	-.09	.06	**.12**	.10	**.34**	-									
8. CS2.3	2.48	1.12	-.10	.08	.03	.04	-.05	**-.23**	-.07	-								
9. CS2.4	3.52	1.09	-.02	-.06	**.20**	.05	**.15**	.08	**.11**	-.01	-							
10. CS2.5	3.27	1.29	.02	-.03	.09	.10	**.19**	**.10**	**.36**	**-.13**	**.21**	-						
11. CS2.6	3.69	1.10	.03	-.01	**.10**	**.15**	.00	**.12**	**.20**	.08	.00	.06	-					
12. HS1.1	2.58	1.34	.00	**.12**	.05	**.11**	.06	.04	-.03	.03	.03	-.06	**.13**	-				
13. HS1.2	1.94	1.02	-.07	**.11**	.00	.09	.02	-.06	-.09	**.34**	-.10	**-.20**	.04	**.16**	-			
14. HS1.3	1.89	.97	**-.15**	.06	.07	**.16**	**.17**	-.04	-.07	.05	.00	-.10	.02	**.20**	**.17**	-		
15. HS1.4	4.21	.71	.07	.01	**.15**	-.07	-.07	.02	.09	**.14**	**.10**	.06	**.21**	-.04	.04	-.06	-	
16. HS1.5	4.05	.92	**.22**	-.07	.05	.01	.05	**.21**	**.24**	**-.13**	**.17**	**.28**	-.01	**-.11**	**-.26**	**-.10**	-.03	-
17. HS1.6	3.66	1.10	**.18**	-.06	.07	.07	.09	.00	**.32**	-.04	.09	**.36**	.03	**-.13**	-.10	**-.12**	-.04	**.28**
18. Helping[a]	19.23	5.62	**.17**	.04	**-.16**	**-.10**	**-.14**	.10	.03	-.05	-.10	-.04	-.01	**-.12**	-.05	**-.18**	**.14**	**.11**
19. HS2.1	3.56	1.16	**.16**	.09	**-.16**	**-.11**	**-.13**	**.11**	.08	-.01	-.04	-.01	-.04	**-.11**	-.06	**-.16**	**.13**	**.14**
20. HS2.2	4.16	.83	**.19**	.05	.06	.01	-.01	.00	-.01	.03	.06	.03	**.12**	.03	.06	-.03	**.13**	.02
21. HS2.3	2.57	1.10	**-.16**	-.04	**.12**	.10	**.15**	-.03	.04	.10	**.17**	.10	-.01	**.12**	.09	**.19**	**-.11**	-.01
22. HS2.4	2.88	1.23	**-.12**	.07	**.10**	.04	.07	-.08	.01	.06	**.10**	.03	-.01	.08	.00	**.13**	**-.10**	-.07
23. HS2.5	2.86	1.10	.00	.06	-.02	.02	-.01	-.03	**.11**	.03	**.16**	**.14**	.01	.05	-.04	-.06	.05	.01

Table 10 (cont'd)

	17. HS1.6	18. Helping	19. HS2.1	20. HS2.2	21. HS2.3	22. HS2.4	23. HS2.5
17. HS1.6	-						
18. Helping[a]	-.09	*.75*					
19. HS2.1	-.06	**.92**	-				
20. HS2.2	.02	**.16**	**.20**	-			
21. HS2.3	.09	**-.76**	**-.57**	-.09	-		
22. HS2.4	.09	**-.82**	**-.59**	-.08	**.53**	-	
23. HS2.5	.03	**-.23**	**-.17**	.01	**.28**	**.18**	-

Table 10 (cont'd)

	M	*SD*	1. CS1.1	2. CS1.2	3. CS1.3	4. CS1.4	5. CS1.5	6. CS2.1	7. CS2.2	8. CS2.3	9. CS2.4	10. CS2.5	11. CS2.6	12. HS1.1	13. HS1.2	14. HS1.3	15. HS1.4	16. HS1.5
24. CS1 SJT score	1.17	1.00	**.17**	**-.11**	**.15**	.03	**-.27**	.00	.01	-.04	.00	.03	.06	.01	-.06	-.05	**.11**	.07
25. CS2 SJT score	.55	.85	.06	-.08	.09	.06	**.11**	**.18**	**.29**	**-.37**	**.14**	**.38**	**.17**	.01	**-.26**	-.04	.01	**.22**
26. HS1 SJT score	.59	.97	-.03	.08	.02	-.02	-.03	**.11**	.00	**-.17**	.09	.03	-.07	.00	**-.32**	.01	**-.11**	**.33**
27. HS2 SJT score	.62	.60	**.19**	.03	-.02	.02	**-.10**	.01	.07	-.06	-.07	-.01	.01	-.06	**-.11**	-.07	.04	**.12**
28. Total SJT score	2.93	1.87	**.16**	-.04	**.12**	.03	**-.14**	**.15**	**.16**	**-.30**	.09	**.20**	.07	-.01	**-.35**	-.06	.02	**.35**
29. Opp Perform	2.94	.93	.06	-.05	.04	**.11**	**.10**	**.14**	**.20**	-.05	.06	.07	.03	-.01	-.03	.00	-.01	**.14**
30. Job Rel. Pre	2.00	.91	-.03	.00	.01	.10	.08	.09	**.13**	.01	.01	.06	.00	.02	.02	.03	.06	.04
31. Job Rel. Con	2.15	.96	.01	-.04	.01	.08	.04	**.11**	.09	-.10	.02	.03	.03	-.03	.01	.09	.00	.06
32. Attraction	3.62	.83	**.22**	-.09	-.05	-.07	.04	**.13**	.07	**-.19**	.01	.08	.09	-.02	-.07	.03	.04	**.12**
33. Compatibility	3.15	.90	**.15**	-.03	-.03	-.02	.06	.05	.05	**-.16**	.02	.08	.06	.00	.00	.04	.07	.06
34. Att_CS1	2.93	.91	.05	.03	-.06	-.03	-.09	-.03	-.02	.02	**-.12**	-.04	.01	.01	-.04	.02	.04	.06
35. Att_CS2	3.12	.98	.04	.01	**-.17**	-.03	-.04	-.07	.02	.04	-.04	-.03	.04	.06	-.02	-.06	.02	-.04
36. Att_HS1	2.97	.99	-.03	**.15**	-.09	-.05	-.06	.02	-.01	.01	-.03	**-.11**	-.02	.01	-.04	.00	.01	-.03
37. Att_HS2	2.92	.98	.06	-.02	-.04	.00	-.03	.01	-.01	-.01	**-.13**	-.04	.00	.01	.04	-.05	.02	.04
38. Acq_CS1.focal (0=No; 1=Yes)	.02	.14	.00	.06	.03	-.01	.03	.01	.04	.05	.02	.08	.06	-.02	.07	.06	.10	.01
39. Acq_CS1.non-focal (0=N; 1=Y)	.03	.16	**-.12**	.01	.07	-.01	.01	-.03	.01	-.04	.07	.00	.00	.03	-.02	.05	-.03	.01
40. Acq_HS1.focal (0=No; 1=Yes)	.02	.14	-.02	-.01	.03	-.06	.00	-.04	.05	.01	-.03	.02	-.03	.00	-.03	.04	.01	-.03
41. Acq_HS1.non-focal (0=N; 1=Y)	.01	.09	**-.11**	.10	.02	-.01	.00	.01	.00	.02	.07	.03	.00	.10	-.02	**.13**	.02	.00
42. Acq_CS2.focal (0=No; 1=Yes)	.02	.15	-.08	.06	-.01	-.03	-.08	-.04	.02	.03	.01	.05	.01	-.02	-.04	.09	-.05	-.03

Table 10 (cont'd)

	17. HS1.6	18. Helping	19. HS2.1	20. HS2.2	21. HS2.3	22. HS2.4	23. HS2.5
24. CS1 SJT score	.02	**.11**	.09	**.13**	**-.12**	-.09	-.05
25. CS2 SJT score	**.23**	-.06	-.07	-.05	.04	.03	.04
26. HS1 SJT score	-.05	**.13**	**.16**	.06	-.06	-.06	.00
27. HS2 SJT score	.09	**.17**	**.18**	**.11**	**-.25**	-.01	-.03
28. Total SJT score	**.12**	**.15**	**.16**	**.11**	**-.15**	-.07	-.02
29. Opp Perform	.02	-.05	.02	.06	**.11**	**.11**	**.12**
30. Job Related Pre	.05	.00	.03	.04	.07	-.01	.04
31. Job Related Con	.03	.02	.06	.01	.08	-.01	.04
32. Attraction	.07	**.11**	**.13**	**.13**	-.04	-.08	-.05
33. Compatibility	.06	.07	.07	.08	.02	-.09	.03
34. Att_CS1	.02	**.12**	**.11**	-.06	**-.15**	-.04	-.04
35. Att_CS2	.01	.06	.07	.02	-.06	-.01	.00
36. Att_HS1	-.03	**.14**	**.16**	.01	**-.12**	-.05	.00
37. Att_HS2	.00	.01	.01	.04	-.04	.03	.02
38. Acq_CS1.focal (0=No; 1=Yes)	-.03	.04	.05	.05	-.02	-.02	.04
39. Acq_CS1.nonfocal (0=No; 1=Yes)	-.04	.00	-.01	-.05	.02	-.02	.07
40. Acq_HS1.focal (0=No; 1=Yes)	-.01	.05	.05	.02	-.05	-.02	.09
41. Acq_HS1.nonfocal (0=No; 1=Yes)	-.05	.07	.06	-.02	-.02	-.09	.04
42. Acq_CS2.focal (0=No; 1=Yes)	-.05	-.02	.00	-.07	.00	.04	.07

Table 10 (cont'd)

	M	*SD*	1. CS1.1	2. CS1.2	3. CS1.3	4. CS1.4	5. CS1.5	6. CS2.1	7. CS2.2	8. CS2.3	9. CS2.4	10. CS2.5	11. CS2.6	12. HS1.1	13. HS1.2	14. HS1.3	15. HS1.4	16. HS1.5
43. Acq_CS2. non-focal (0=N; 1=Y)	.01	.12	.02	.08	.01	.00	-.04	-.01	-.03	.08	-.03	.01	-.01	-.05	-.02	-.01	.00	-.03
44. Acq_HS2. focal (0=N; 1=Y)	.01	.07	-.07	**.10**	.00	-.02	.00	.04	-.01	.03	.07	.04	.02	.05	-.03	.08	.03	.00
45. Acq_HS2.non-focal (0=N; 1=Y)	.03	.18	-.01	.03	-.02	.00	-.04	-.08	-.06	.05	-.02	-.01	-.09	-.06	.04	-.04	-.06	-.04
46. # Actor(s) knew	.15	.63	-.08	.08	.03	-.03	-.03	-.05	.00	.04	.02	.04	-.02	-.01	-.01	.07	-.01	-.03
47. Gender (0=Female; 1=Male)	.49	.50	-.03	.01	-.03	.03	-.05	-.08	**-.13**	-.01	.00	-.05	.05	.06	.06	**.13**	-.06	-.09
48. Race[b] (0=Minority; 1=White)	.78	.42	.05	-.06	**-.15**	-.09	-.03	.06	.03	.04	-.07	-.04	.02	-.10	-.07	-.03	.02	-.02
49. Age	19.97	2.68	.01	-.02	-.03	-.07	.01	-.06	-.08	-.04	**-.10**	.04	-.05	-.06	-.01	-.05	-.07	-.04
50. Year in college	2.39	1.08	-.03	-.01	-.07	-.09	-.02	-.06	-.02	.00	-.06	-.01	.04	.03	.00	-.02	-.05	-.04
51. GPA[c]	5.95	1.89	**.13**	-.01	-.05	.01	-.02	.08	.09	.09	.00	.01	.08	-.06	.02	-.05	.04	.09

Table 10 (cont'd)

	17. HS1.6	18. Helping	19. HS2.1	20. HS2.2	21. HS2.3	22. HS2.4	23. HS2.5
43. Acq_CS2.nonfocal (0=No; 1=Yes)	.06	-.01	-.02	-.02	-.04	.03	.06
44. Acq_HS2.focal (0=No; 1=Yes)	-.04	.05	.06	.03	.00	-.05	.04
45. Acq_HS2.nonfocal (0=No; 1=Yes)	.03	-.05	-.02	.00	.05	.08	.06
46. # Actor(s) knew	-.02	.02	.03	-.02	-.01	.01	.10
47. Gender (0=Female; 1=Male)	-.09	-.06	**-.11**	-.10	.01	-.01	**-.18**
48. Race[b] (0=Minority; 1=White)	.05	.05	.04	.07	-.02	-.05	.00
49. Age	.05	.01	-.04	-.01	-.08	-.03	**-.13**
50. Year in college	.01	-.02	-.06	-.04	-.05	-.01	**-.13**
51. GPA[c]	.05	**.16**	**.15**	**.12**	**-.11**	**-.13**	-.01

Table 10 (cont'd)

	24. CS1 SJT score	25. CS2 SJT score	26. HS1 SJT score	27. HS2 SJT score	28. Total SJT score	29. Opp Perform	30. Job Rel. Pre	31. Job Rel. Con	32. Attraction	33. Compat-ibility	34. Att_CS1	35. Att_CS2	36. Att_HS1	37. Att_HS2	38. Acq_CS1. focal	39. Acq_CS1. nonfocal	40. Acq_HS1. focal	41. Acq_HS1. nonfocal
24. CS1 SJT score	-																	
25. CS2 SJT score	.03	-																
26. HS1 SJT score	.03	.06	-															
27. HS2 SJT score	.08	.07	.08	-														
28. Total SJT score	**.59**	**.53**	**.59**	**.44**	-													
29. Opp Perform	.07	**.17**	.07	.01	**.15**	*.84*												
30. Job Rel. Pre	-.02	.02	.00	-.02	-.01	**.48**	*.81*											
31. Job Rel. Con	-.03	**.10**	.07	-.01	.06	**.39**	**.54**	*.83*										
32. Attraction	**.11**	**.13**	**.10**	**.17**	**.23**	**.14**	**.15**	**.15**	*.88*									
33. Compatibility	.07	**.12**	**.11**	**.10**	**.18**	**.14**	**.15**	**.11**	**.75**	*.85*								
34. Att_CS1	.06	.01	-.01	.08	.05	.00	-.01	-.04	.06	-.06	-							
35. Att_CS2	-.03	-.01	-.01	-.01	-.03	.05	.06	.02	**.16**	.08	**.36**	-						
36. Att_HS1	.01	**-.10**	.07	-.03	-.02	.06	-.01	.03	.04	-.04	**.48**	**.35**	-					
37. Att_HS2	-.02	-.09	.03	.02	-.03	.06	-.02	.02	.05	.04	**.23**	**.23**	**.32**	-				
38. Acq_CS1.focal (0=N; 1=Y)	-.04	.03	-.04	-.04	-.05	**.12**	.03	.02	.06	.00	.01	.06	.06	.03	-			
39. Acq_CS1.non-focal (0=N; 1=Y)	.00	**.11**	.02	.02	.07	.01	-.07	.04	.05	.07	.00	.10	.04	-.02	**.22**	-		
40. Acq_HS1.focal (0=N; 1=Y)	.00	-.07	-.06	.05	-.05	-.08	-.02	-.04	-.06	.00	.06	.08	.00	.01	**.13**	**.22**	-	
41. Acq_HS1.non-focal (0=N; 1=Y)	.01	-.02	-.02	-.04	-.03	.04	.02	.03	.04	.03	-.03	.08	.09	-.05	**.43**	**.36**	**.43**	-

Table 10 (cont'd)

	24. CS1 SJT score	25. CS2 SJT score	26. HS1 SJT score	27. HS2 SJT score	28. Total SJT score	29. Opp Perform	30. Job Rel. Pre	31. Job Rel. Con	32. Attraction	33. Compat-ibility	34. Att_CS1	35. Att_CS2	36. Att_HS1	37. Att_HS2	38. Acq_CS1. focal	39. Acq_CS1. nonfocal	40. Acq_HS1. focal
42. Acq_CS2. focal (0=N; 1=Y)	-.10	-.04	-.02	.04	-.07	.00	.01	.01	-.03	-.06	-.04	.00	.00	-.04	**.37**	.08	**.37**
43. Acq_CS2.non-focal (0=N; 1=Y)	.00	-.02	-.07	.07	-.02	-.05	-.01	.02	.07	.08	-.04	.06	-.07	-.04	**.16**	**.27**	**.16**
44. Acq_HS2. focal (0=N; 1=Y)	-.01	.04	-.01	-.01	.00	.00	-.06	.01	.05	.02	-.03	**.10**	**.11**	-.07	**.53**	**.44**	**.26**
45. Acq_HS2.non-focal (0=N; 1=Y)	.04	-.07	-.09	.07	-.03	**.15**	.05	.05	.02	.04	-.03	-.02	.02	.08	**.19**	**.24**	**.19**
46. # Actor(s) knew	-.02	-.01	-.07	.04	-.04	.05	-.01	.03	.04	.03	-.02	.08	.05	-.01	**.60**	**.59**	**.56**
47. Gender (0=Female; 1=Male)	.03	.02	-.07	-.08	-.03	.00	-.02	.08	.09	**.17**	**-.17**	**-.17**	**-.13**	**-.13**	-.02	.03	-.02
48. Race[b] (0=Minority; 1=White)	.04	.01	.02	**.14**	.08	**-.13**	**-.14**	**-.13**	.00	.04	-.01	.06	.06	-.03	**-.12**	**-.16**	-.02
49. Age	.03	.02	**-.13**	-.04	-.06	-.06	-.01	.00	.02	-.01	-.05	.00	-.02	.07	.01	.01	.08
50. Year in college	-.06	-.01	**-.16**	-.05	**-.14**	**-.16**	-.06	-.08	.04	.03	-.01	.02	-.05	.10	.00	-.03	.10
51. GPA[c]	.03	.01	.06	.09	.08	.03	.04	.05	-.05	**-.11**	.03	.03	.02	-.07	.05	-.03	.06

Table 10 (cont'd)

	41. Acq_HS1.non-focal	42. Acq_CS2.focal	43. Acq_CS2.non-focal	44. Acq_HS2.focal	45. Acq_HS2.non-focal	46. # Actor(s) knew	47. Gender (0,1)	48. Race (0,1)	49. Age	50. Year in college	51. GPA
42. Acq_CS2.focal (0=No; 1=Yes)	**.38**	-									
43. Acq_CS2.nonfocal (0=No; 1=Yes)	**.25**	**.13**	-								
44. Acq_HS2.focal (0=No; 1=Yes)	**.82**	**.23**	**.31**	-							
45. Acq_HS2.nonfocal (0=No; 1=Yes)	**.31**	**.16**	**.36**	**.19**	-						
46. # Actor(s) knew	**.74**	**.57**	**.53**	**.68**	**.61**	-					
47. Gender (0=Female; 1=Male)	.09	.09	.02	.07	-.01	.05	-				
48. Race[b] (0=Minority; 1=White)	-.10	.00	-.01	-.05	-.09	**-.12**	-.05	-			
49. Age	.03	.02	.03	-.01	.00	.04	**.20**	**-.11**	-		
50. Year in college	-.03	-.02	.02	-.06	.01	.00	**.12**	-.09	**.55**	-	
51. GPA[c]	.00	.05	**-.11**	.02	-.10	-.02	**-.12**	**.19**	**-.14**	**-.17**	-

Note. Correlations in bold are significant at at least the .05 level (2-tailed). *N*s range from 348 (for race) to 374. CS1= conflict situation 1. CS2= conflict situation 2. HS1= helping situation 1. HS2= helping situation 2. Att_CS1, Att_CS2, Att_HS1, and Att_HS2 refer to the rated attractiveness of the non-focal actor in conflict situation 1, conflict situation 2, helping situation 1, and helping situation 2, respectively. Variables starting with "Acq" refer to whether or not the respondent was acquainted with the focal or non-focal actor in a particular SJT item. [a]Helping [score] is composed of HS2.1, HS2.3, and HS2.4. [b]Race variable excludes those who were multiracial or "other." [c]GPA was a categorical variable.

Tests of Hypotheses Pertaining to Helping

Before discussing the hypotheses pertaining to helping situations, I include Tables 11 and 12 below to provide an overview of mean likelihood ratings as a function of sex (Table 11) and race (Table 12) of the respondent and non-focal actor.

Table 11. **Likelihood ratings for helping response options as a function of respondent's and non-focal actor's sex**

	Female Respondent				Male Respondent			
	Female non-focal actor (*N* = 96)		Male non-focal actor (*N* = 93)		Female non-focal actor (*N* = 95)		Male non-focal actor (*N* = 90)	
Response Option	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HS1.1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.	2.69	1.48	2.31	1.24	2.87	1.28	2.42	1.27
HS1.2: Suggest that your coworker ignore Frank because working with Frank can be difficult.	2.01	1.05	1.74	1.01	2.05	1.00	1.93	.99
HS1.3: Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank.	1.74	1.00	1.80	.85	2.04	1.08	1.98	.92
HS1.4: Tell your coworker how you deal with Frank.	4.24	.56	4.27	.74	4.26	.64	4.08	.86
HS1.5: Suggest that your coworker meet with Frank to talk about their differences.	4.17	.85	4.09	.88	4.04	.96	3.88	.98
HS1.6: Suggest that your coworker talk to the team leader about his/her problem with Frank.	3.77	1.11	3.74	1.01	3.54	1.14	3.59	1.13
HS2.1: Agree to switch with your coworker so that he/she can take the training course.	3.69	1.14	3.69	1.08	3.67	1.15	3.17	1.20
HS2.2: Agree to switch with your coworker, but only if he/she will switch with you in a few weeks.	4.15	.85	4.33	.86	4.18	.68	3.96	.90
HS2.3: Tell your coworker you think he/she is taking advantage of the system and you won't switch.	2.52	1.10	2.59	1.13	2.43	1.05	2.74	1.11
HS2.4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	2.81	1.23	2.96	1.36	2.65	1.16	3.10	1.13
HS2.5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	3.06	1.00	3.05	1.11	2.46	1.09	2.88	1.10

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

Table 12. **Likelihood ratings for helping response options as a function of respondent's and non-focal actor's race**

	Asian Respondent						Black Respondent						White Respondent					
	Asian non-focal actor (*N* = 17)		Black non-focal actor (*N* = 12)		White non-focal actor (*N* = 12)		Asian non-focal actor (*N* = 11)		Black non-focal actor (*N* = 14)		White non-focal actor (*N* = 10)		Asian non-focal actor (*N* = 94)		Black non-focal actor (*N* = 89)		White non-focal actor (*N* = 88)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HS1.1	3.47	1.33	3.00	1.65	2.67	1.30	3.36	1.36	2.29	1.07	2.20	1.03	2.34	1.30	2.62	1.34	2.66	1.37
HS1.2	2.12	.93	2.75	1.29	1.92	1.00	2.27	1.10	1.57	.85	1.70	1.06	1.94	1.01	1.93	.99	1.86	1.01
HS1.3	2.35	1.37	1.58	.67	2.42	1.08	1.91	1.22	1.64	1.01	1.50	.97	1.66	.78	1.85	.94	2.11	1.01
HS1.4	4.06	.75	4.25	.87	4.67	.49	4.00	.45	4.07	.73	4.10	.57	4.20	.61	4.33	.65	4.15	.84
HS1.5	4.12	.93	3.75	1.06	3.92	1.00	4.09	.94	4.50	.65	4.00	.94	4.16	.79	3.98	.98	3.98	.97
HS1.6	3.65	.93	3.42	.90	2.92	1.51	3.82	1.08	3.71	1.07	3.80	1.14	3.83	1.09	3.70	1.10	3.52	1.12
HS2.1	3.29	1.16	3.33	1.30	4.08	1.08	2.45	1.04	3.93	.92	3.80	.92	3.51	1.10	3.48	1.24	3.81	1.07
HS2.2	4.24	.97	4.17	.72	3.83	1.19	3.91	.94	4.21	.70	3.80	1.14	4.17	.80	4.27	.75	4.16	.80
HS2.3	2.53	1.12	2.42	1.56	2.17	.94	3.36	1.03	2.36	1.22	2.90	.88	2.53	1.04	2.62	1.02	2.41	1.14
HS2.4	3.00	1.27	3.00	1.28	2.17	1.19	3.73	.79	3.21	1.48	2.90	1.29	2.83	1.26	2.82	1.15	2.83	1.23
HS2.5	2.94	1.03	2.67	1.30	2.33	.98	3.00	1.00	2.79	1.19	3.60	1.07	2.91	1.11	2.83	1.04	2.81	1.08

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

H1, scoring method 1. Hypothesis 1 stated that respondent race will moderate the effects of the race mismatch between the respondent and the non-focal actor on helping, such that the effects of race mismatch will be more negative for White respondents viewing a minority non-focal actor than for minority respondents viewing a White non-focal actor. To test this hypothesis, an analysis of variance (ANOVA) was conducted with dichotomized respondent race (White, minority) and race match as the independent variables. The ANOVA showed no significant effect of the race by race match interaction on helping, $F(1,316) = .02$, $MSE = 31.44$, ns. Thus, Hypothesis 1 was not supported using scoring method 1.

H1, scoring method 2. Additionally, a MANOVA was calculated to examine the effects of race mismatch and respondent race on likelihood ratings respondents provided for the eleven helping response options. These were entered in as a set of dependent variables into the MANOVA. There was no significant effect of the respondent race by race match interaction on likelihood ratings on the set of response options, Wilks' lambda = .96, $F(11,296) = 1.13$, ns. Thus, Hypothesis 1 was not supported using scoring method 2.

H1, scoring method 3. Scoring method 3 was applied to examine the effects of race mismatch and respondent race on effectiveness of responses to helping situations 1 and 2. The interaction of race mismatch and dichotomized respondent race on helping situation 1 score was not statistically significant, $F(1,329) = .97$, $MSE = .92$, ns. The main effect of race mismatch was marginally significant, $F(1,329) = 3.76$, $MSE = .92$, $p = .053$. Respondents scored higher on helping situation 1 when the non-focal actor was of the same race as them ($M = .71$, $SD = .94$) as opposed to when there was a race mismatch ($M = .53$, $SD = .96$).

Next, looking at helping situation 2, the interaction of race mismatch and dichotomized respondent race on helping situation 2 score was significant, $F(1,316) = 9.52$, $MSE = .33$, $p <$

.01, as were the main effects for race mismatch, $F(1,316) = 9.13$, $MSE = .33$, $p < .01$, and for dichotomized respondent race, $F(1,316) = 6.39$, $p < .05$. White respondents scored the same on helping item 2 regardless of whether there was a race match ($M = .68$, $SD = .52$) or a race mismatch ($M = .68$, $SD = .59$) with the non-focal actor, $F(1,250) = .01$, $MSE = .32$, ns. Minority respondents, however, scored higher on average on helping item 2 when the non-focal actor was of the same race as them ($M = .77$, $SD = .65$) than when there was a race mismatch between the respondent and non-focal actor ($M = .23$, $SD = .58$), $F(1, 66) = 10.81$, $MSE = .37$, $p < .01$. Thus, with respect to response effectiveness, there were (marginally) negative effects of race mismatch on both minority and White respondents for helping situation 1 and more negative effects of race mismatch on minority than on White respondents for helping situation 2.

To find a substantive explanation for the significant effect of race mismatch on minority respondents' helping situation 2 scores, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses with the goal of examining how the frequencies with which different options were selected as most or least likely were affected by race match with the non-focal actor. Race match did not have a statistically significant effect on the frequency with which various response options were selected as most likely, $\chi^2(3, N = 333) = 5.51$, ns, or least likely, $\chi^2(3, N = 333) = 4.97$, ns. Therefore, actual frequencies of responses were not further examined.

H2, scoring method 1. To test Hypothesis 2, that there will be a main effect for respondent sex on helping, such that female respondents will offer more help to non-focal actors than will male respondents, an ANOVA was performed. The effect of gender was not significant, $F(1,345) = 1.51$, $MSE = 31.73$, ns. Thus, Hypothesis 2 was not supported using scoring method 1.

H2, scoring method 2. Additionally, a MANOVA was calculated to examine the effect of respondent sex on likelihood ratings respondents provided for the entire set of helping response options. The MANOVA showed a significant effect of respondent sex on the set of dependent variables, Wilks' lambda = .93, $F(11,325) = 2.39$, $p < .01$. Follow up analyses showed that female respondents gave significantly lower likelihood ratings ($M = 1.79$, $SD = .94$) than male respondents ($M = 2.02$, $SD = .99$) to helping situation 1's response option 3 (Tell your coworker that he/she needs to learn how to handle his/her own conflicts with Frank), $t(335) = -2.18$, $p < .05$. Response option 3 was rated by judges as low on helping ($M = 1.30$ on 3-point scale, $SD = .61$), indicating that women's lower average likelihood rating for this option can be said to correspond to a higher willingness (relative to male respondents) to help the non-focal actor. Female respondents also gave significantly higher likelihood ratings ($M = 3.03$, $SD = 1.06$) than male respondents ($M = 2.64$, $SD = 1.12$) to helping situation 2's response option 5 (Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch), $t(335) = 3.30$, $p = .001$. Response option 5 was rated by judges as above average on helping ($M = 2.33$ on a 3-point scale, $SD = .86$), suggesting that women's higher average likelihood rating for this option (relative to men's) should correspond to a higher willingness to help the non-focal actor. Finally, women also gave somewhat higher likelihood ratings than men to option 6 of helping situation 1 (Suggest that your coworker talk to the team leader about his/her problem with Frank), $t(335) = 1.81$, $p = .072$, and option 2 of helping situation 2 (Agree to switch with your coworker, but only if he/she will switch with you in a few weeks), $t(335) = 1.91$, $p = .057$. Female respondents' higher average ratings on these options ($M = 3.74$ with $SD = 1.09$ and $M = 4.26$ with $SD = .85$, respectively) relative to male respondents' ($M = 3.52$ with $SD = 1.15$ and $M = 4.09$ with $SD = .80$, respectively) should correspond to a

higher willingness to help the non-focal actor given that judges rated these two response options highly on helping (M = 2.5 with SD = .72 and M = 2.95 with SD = .22, respectively). Thus, Hypothesis 2 was partially supported using scoring method 2.

H2, scoring method 3. Next, scoring method 3 was applied to examine the effects of respondent sex on effectiveness of responses to helping situations 1 and 2. There was no statistically significant main effect for respondent sex on helping situation 1 score, $F(1,360)$ = 1.42, MSE = .94, ns. The main effect of respondent sex on helping situation 2 score was marginally significant, $F(1,345)$ = 2.85, MSE = .36, p = .092. Women scored somewhat higher (M = .68, SD = .59) on helping situation 2 than men (M = .57, SD = .61). Thus, with respect to response effectiveness, there were no significant differences between male and female respondents.

H3, scoring method 1. Hypothesis 3 stated that respondent sex will moderate the effects of non-focal actor sex on helping, such that a female non-focal actor will receive more help than a male non-focal actor when the respondent is male while female and male non-focal actors will receive equal amounts of help when the respondent is female. An ANOVA showed a significant effect of the non-focal actor sex by respondent sex interaction on helping, $F(1,343)$ = 3.93, MSE = 30.98, $p < .05$. The main effect for non-focal actor sex was also significant, $F(1,343)$ = 6.29, MSE = 30.98, $p < .05$. To break down the significant two way interaction, separate ANOVAs for male and female respondents were calculated and revealed a statistically significant simple main effect of non-focal actor sex for male respondents, $F(1,173)$ = 10.62, MSE = 29.61, p = .001, but not for female respondents, $F(1,170)$ = .13, MSE = 32.39, ns. Male respondents were willing to offer more help on average to female (M = 20.17, SD = 5.28) than to male coworkers (M = 17.49, SD = 5.61). Female respondents were willing to offer equal help to female (M = 19.81, SD

= 5.58) and male coworkers (*M* = 19.50, *SD* = 5.80). Thus, Hypothesis 3 was supported using scoring method 1.

H3, scoring method 2. As an additional test of Hypothesis 3 at the level of the helping response options, a MANOVA was calculated with non-focal actor sex and respondent sex as independent variables. The interaction of the two independent variables was not statistically significant, Wilks' lambda = .95, $F(11, 323) = 1.43$, ns. This means that Hypothesis 3 did not receive support at the level of the individual response options. Consistent with the earlier analyses with regard to Hypothesis 2 though, there was a significant main effect for respondent gender, Wilks' lambda = .93, $F(11,323) = 2.37$, $p < .01$. The main effect for non-focal actor gender was also significant, Wilks' lambda = .93, $F(11,323) = 2.28$, $p < .05$. On average, respondents gave helping situation one's options 1 and 2 higher likelihood ratings when the non-focal actor was female, $t(359.06) = 2.59$, $p = .01$, and $t(360) = 1.91$, $p = .057$, respectively. Respondents also gave lower ratings to helping situation two's response options 4 and 5 when the non-focal actor was female, $t(345) = -2.57$, $p < .05$, and $t(345) = -1.90$, $p = .058$, respectively. The means and standard deviations by non-focal actor gender for likelihood ratings on these three response options are shown in Table 13. Response option 1 (Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank) is on the higher end of helping (*M* = 2.30 on a 3-point scale, *SD* = .85), response option 2 (Suggest that your coworker ignore Frank because working with Frank can be difficult) is on the lower end (*M* = 1.43 on a 3-point scale, *SD* = .75), response option 4 (Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule) is on the lower end (*M* = 1.63 on a 3-point scale, *SD* = .81) and option 5 (Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her

find someone else to switch) is on the higher end (*M* = 2.33 on a 3-point scale, *SD* = .86). In the context of the mean differences in likelihood ratings this suggests that respondents were more willing to help female non-focal actors in the case of response options 1 and 4 but were somewhat more willing to help male non-focal actors in the cases of response options 2 and 5.

Table 13. **Likelihood ratings for three response options as a function of non-focal actor sex**

	Female non-focal actor			Male non-focal actor		
Rated Response Option	*M*	*SD*	*N*	*M*	*SD*	*N*
HS1 Op1: Explain to your coworker that you are very busy, but if he/she wants to talk later, you'll tell him/her how you handle Frank.	2.75	1.39	185	2.39	1.26	177
HS1 Op2: Suggest that your coworker ignore Frank because working with Frank can be difficult.	2.02	1.03	185	1.82	.99	177
HS2 Op4: Tell your coworker you won't switch, and he/she should wait until training is offered at another time that doesn't interfere with his/her schedule.	2.69	1.18	180	3.03	1.25	167
HS2 Op5: Tell your coworker that you think he/she is taking advantage of the system, but you will help him/her find someone else to switch.	2.73	1.09	180	2.95	1.10	167

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

H3, scoring method 3. Next, scoring method 3 was applied to examine the effects of non-focal actor sex and respondent sex on effectiveness of responses to helping situations 1 and 2. The interaction of non-focal actor sex and respondent sex on helping situation 1 score and on helping situation 2 score was not statistically significant in both cases, $F(1,358) = .37$, $MSE = .94$, ns, and $F(1,343) = .52$, $MSE = .36$, ns, respectively. Consistent with analyses reported earlier, there was a marginally significant main effect of respondent sex on helping situation 2

score, $F(1,343) = 2.99$, $MSE = .36$, $p = .085$, with women scoring slightly higher than men on the item.

H4, scoring method 2. Hypothesis 4, which stated that in a helping situation where the competence stereotype is activated, the lowest level of help will be offered by White male respondents to male Black non-focal actors as compared to other race and sex-based combinations of respondents with targets, could not be tested as planned using scoring method 1. Method 2 was used to do the analysis at the level of the response options associated with helping situation 1 (the relevant situation for this hypothesis). A MANOVA with respondent sex (male, female), dichotomized respondent race (minority, White), non-focal actor sex (male, female), and non-focal actor race (Black, White) as the independent variables and likelihood ratings for the six response options associated with helping situation 1 as the dependent variables was calculated. Table 14 displays the 16 groups that result from the four way interaction of the independent variables along with the associated sample sizes. Notably, Asian non-focal actors were not included in this analysis. Consistent with earlier analyses, there was a main effect for respondent gender, Wilks' lambda = .91, $F(6,198) = 3.26$, $p < .01$. The MANOVA did not show a statistically significant four-way interaction, Wilks' lambda = .99, $F(6,198) = .34$, ns. The results at the response option level, therefore, do not lend support to Hypothesis 4. It should be noted though that given sample size limitations the power for finding the effect was very low.

Table 14. **Groups compared in testing Hypothesis 4**

Respondent Gender	Respondent Race	Non-focal Actor Gender	Non-focal Actor Race	*N*
Female	Minority (Asian, Black)	Female coworker	Black coworker	5
			White coworker	5
		Male coworker	Black coworker	7
			White coworker	5
	White	Female coworker	Black coworker	26
			White coworker	21
		Male coworker	Black coworker	20
			White coworker	21
Male	Minority (Asian, Black)	Female coworker	Black coworker	11
			White coworker	4
		Male coworker	Black coworker	3
			White coworker	9
	White	Female coworker	Black coworker	19
			White coworker	23
		Male coworker	Black coworker	23
			White coworker	17

H4, scoring method 3. Next, scoring method 3 was applied to examine the effects of the four way interaction of respondent sex, dichotomized respondent race, non-focal actor sex, and non-focal actor race on effectiveness of responses to helping situation 1. The four-way interaction was not statistically significant, $F(1,203) = 1.19$, $MSE = .87$, ns. Thus, the interaction of respondent sex, respondent race, non-focal actor sex, and non-focal actor race did not significantly impact how effectively respondents answered helping item 1. Again, the power for finding an effect was low. What was statistically significant though was the three way interaction between respondent sex, non-focal actor sex, and non-focal actor race (Black, White), $F(1,203) = 4.34$, $MSE = .87$, $p < .05$. Separate ANOVAs were done by non-focal actor race. With a Black non-focal actor, men and women did not score differently depending on whether the Black actor was male or female, $F(1,110) = .13$, $MSE = .88$, ns. With a White non-focal actor, however, men

and women scored differently on helping item 1 depending on the gender of the actor, $F(1,101) = 4.39$, $MSE = .82$, $p < .05$. Male respondents ($M = .74$, $SD = .90$) and female respondents ($M = .69$, $SD = .97$) got equal scores on average on helping item 1 when the White non-focal actor was female. $F(1,51) = .04$, $MSE = .88$, ns, but when the White non-focal actor was male, female respondents ($M = .96$, $SD = .87$) scored higher on the item than male respondents ($M = .27$, $SD = .87$), $F(1,50) = 8.18$, $MSE = .76$, $p < .01$. These differences are displayed in Figure 20 below.

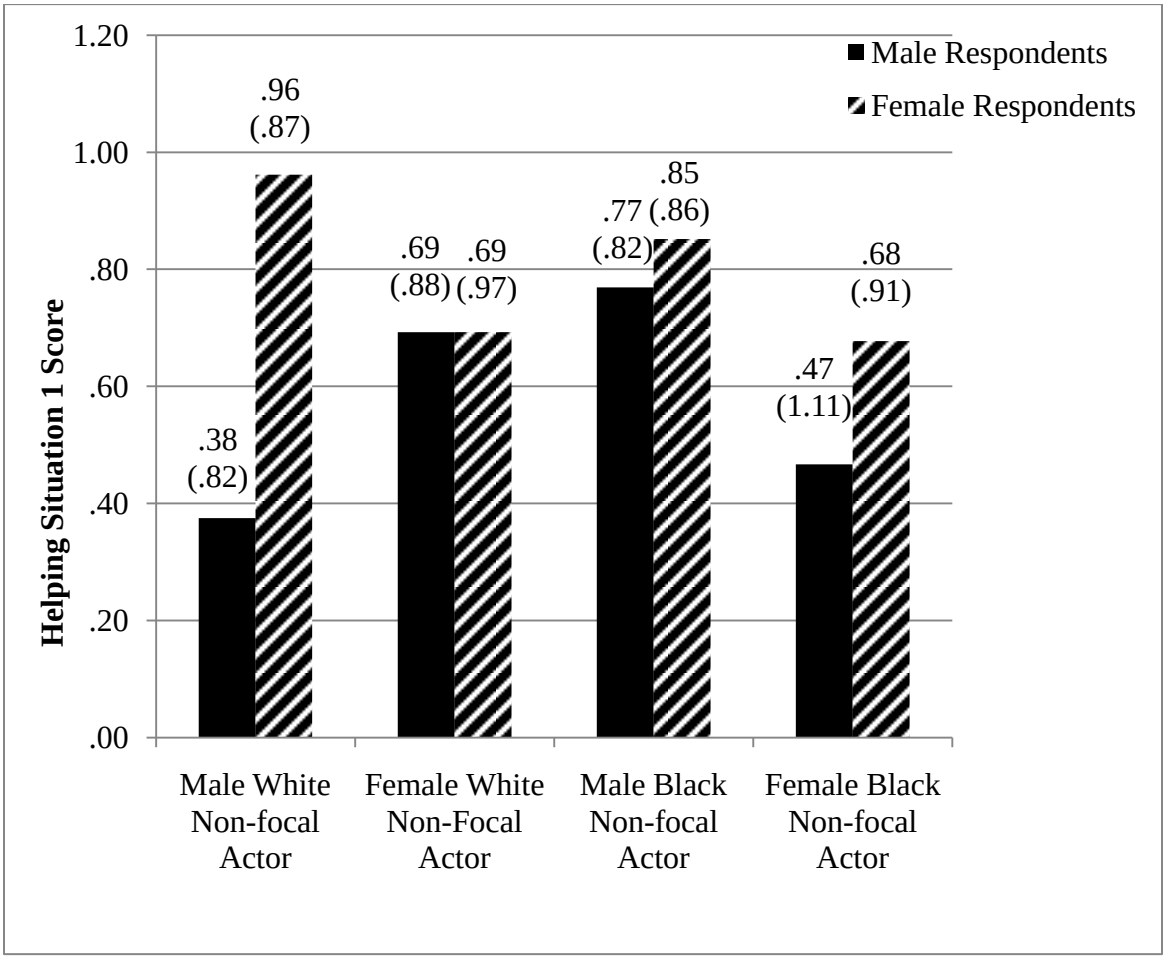

Figure 20. **Effects of non-focal actor race and gender on helping situation 1 score for male and female respondents.**
Standard deviations are below the group means in parentheses.

To try to find a substantive explanation for women's advantage on helping item 1 relative to men when the non-focal actor was a White male, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses (to examine how the frequencies with which different options were selected as most or least likely were affected by respondent gender). Respondent gender did not have a statistically significant effect on the frequency with which various response options were selected as most likely, $\chi^2(5, N = 52) = 8.03$, ns, or least likely, $\chi^2(3, N = 52) = 4.25$, ns. Actual frequencies of responses were not further examined.

H5a, scoring method 1. Hypothesis 5a predicted that in a helping situation where the work ethic stereotype is activated, the lowest level of help will be offered by White respondents to Black non-focal actors as compared to other race-based combinations of respondents with targets. Scoring method 1 was applied to test this hypothesis at the level of the helping dimension. Respondent race (Asian, Black, White) and non-focal actor race (Asian, Black, White) were entered as independent variables into an ANOVA. The main effect for non-focal actor race was significant, $F(2,313) = 3.41$, $MSE = 30.56$, $p < .05$. There was only a marginally significant interaction of respondent and non-focal actor race on helping score, $F(4,313) = 2.01$, $MSE = 30.56$, $p = .094$. Thus, Hypothesis 5a was not supported using scoring method 1.

H5a, scoring method 2. Scoring method 2 was applied next in order to evaluate Hypothesis 5a at the level of the response options associated with helping situation 2. A MANOVA was calculated with race of respondent and race of non-focal actor as the independent variables. The interaction effect on likelihood ratings given to the five response options as a set was not statistically significant, Wilks' lambda = .92, $F(20, 1025.79) = 1.33$, ns. Thus, Hypothesis 5a was also not supported at the level of these individual response options. The main effect for non-focal actor race was significant though, Wilks' lambda = .96, $F(10,618) = 1.97$, p

< .05, but a follow up ANOVA with post hoc Tukey tests on response option 1(for which the main effect of non-focal actor race had been significant) did not show significant pairwise differences between likelihood ratings by non-focal race.

H5a, scoring method 3. Next, scoring method 3 was applied to examine the effects of the interaction of respondent race and non-focal actor race on effectiveness of responses to helping situation 2. The main effect of respondent race and the interaction of respondent with non-focal actor race were statistically significant, $F(2,313) = 5.35$, $MSE = .34$, $p < .01$, and $F(4,313) = 2.91$, $MSE = .34$, $p < .05$, respectively. To break down the interaction separate ANOVAs were done by respondent race. Helping item 2 scores for Asian respondents and White respondents did not differ depending on the race of the non-focal actor, $F(2,35) = 1.05$, $MSE = .36$, ns, and $F(2,250) = .16$, $MSE = .32$, ns, respectively. Black respondents, on the other hand, answered the item less effectively (again, as defined by the test developer) when the non-focal actor was Asian ($M = .00$, $SD = .63$) than when the non-focal actor was Black ($M = .77$, $SD = .73$), $F(2,28) = 4.32$, $MSE = .42$, $p < .05$. (And the difference between the Black and White non-focal actor conditions was not significant as indicated by a post hoc Tukey test.) These results are shown in Figure 21 below.

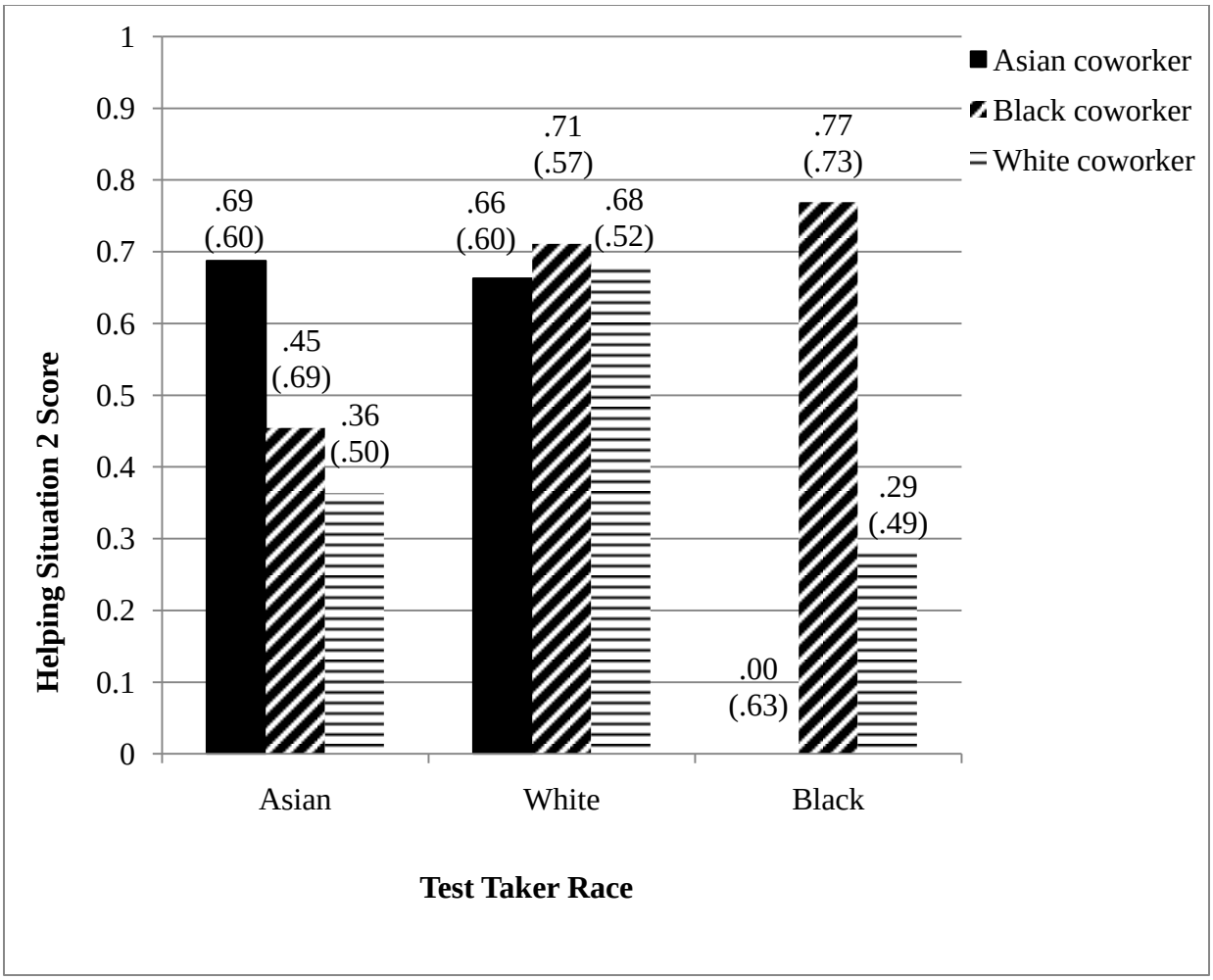

Figure 21. **Effects of non-focal actor race on helping situation 2 score for different groups of respondents.**
Standard deviations are below the group means in parentheses.

To try to find a substantive explanation for the significant effect of non-focal actor race on Black respondents' helping situation 2 scores, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses. Non-focal actor race (Asian, Black) did not have a statistically significant effect on the frequency with which various response options were selected as most likely, $\chi^2(4, N = 24) = 3.39$, ns (Fisher's exact test likewise did not support rejecting the null hypothesis of no effect for non-focal actor race). The effect on least

likely responses, however, was significant, $\chi^2(4, N = 24) = 11.12$, $p < .05$ (Fisher's exact test likewise supported rejecting the null hypothesis of no effect). Therefore, actual frequencies of responses could be further examined. Table 15 shows that a greater proportion of Black respondents in the higher scoring Black non-focal actor group than in the lower scoring Asian non-focal actor group tended to pick the "correct" response option (3; Tell your coworker you think he/she is taking advantage of the system and you won't switch) as their least likely response to helping item 2. For the lower scoring group, the biggest distracter from the correct response appeared to be response option 1 (Agree to switch with your coworker so that he/she can take the training course; scored as a "0"). Thinking in terms of how judges rated each of these options on level of helping, the correct *least likely* response option—3—was the least helpful one ($M = 1.33$ on a 3-point scale, $SD = .66$), while option 1 was one of the most helpful ones ($M = 2.70$ on a 3-point scale, $SD = .65$). While members of the higher scoring group tended to indicate that acting in an unhelpful manner was the least likely thing that they would do if asked to switch shifts with a coworker, members of the lower scoring group were not as good at recognizing and/or showing that this would be the worst way to respond to the situation and were more likely to indicate instead that one of the more helpful options was something they would be least likely to do in the situation. The implication based on this analysis is that for helping scenario 2, Black respondents were more willing to be helpful to a non-focal actor of the same race than to an Asian non-focal actor. It is interesting to note, though, that this effect did not emerge in the above reviewed analyses using likelihood ratings assigned to helping response options as depending variables (H5a, scoring methods 1 and 2).

Table 15. **Observed frequencies of Black respondents' least likely responses on helping situation 2**

	Helping Situation 2 Response Options					
	1	2	3[a]	4	5	Row Total
Asian Non-focal Actor	6	1	1	1	2	11
	54.5%	9.1%	9.1%	9.1%	18.2%	100%
Black Non-focal Actor	1	0	7	4	1	13
	7.7%	0.0%	53.8%	30.8%	7.7%	100%

Note .[a]Considered effective response.

H5b, scoring method 1. To test Hypothesis 5b, that in a helping situation where the work ethic stereotype is activated, the highest level of help will be offered to Asian non-focal actors, an ANOVA with non-focal actor's race as the independent variable was computed. The ANOVA showed no statistically significant effect of non-focal actor's race on respondent willingness to help, $F(2,344) = 1.45$, $MSE = 31.69$, ns. Hypotheses 5b was therefore not supported at the level of the helping dimension (i.e., using scoring method 1).

H5b, scoring method 2. Next, scoring method 2 was applied to calculate a MANOVA on the response options associated with helping situation 2. Hypothesis 5b was also not supported at the level of these individual response options as the main effect for non-focal actor race on likelihood ratings respondents assigned to the two response options was not significant, Wilks' lambda = .97, $F(10,680) = .92$, ns.

H5b, scoring method 3. Next, scoring method 3 was applied to examine the effects of non-focal actor race on effectiveness of responses to helping situation 2. There was no significant main effect of non-focal actor race on helping item 2 score, $F(2,344) = .43$, $MSE = .36$, ns. This lack of a main effect on the second helping item is consistent with earlier analyses

using this same scoring method 3, which instead of a main effect show non-focal actor race to interact with respondent race in determining how effectively respondents answer the item.

Next, results pertaining to Hypotheses 6 through 16 (about conflict) are reviewed. Before proceeding, results for Hypotheses 1 through 5 (about helping) are summarized in Table 16 below. Also, the results of the analyses pertaining to response *effectiveness* are summarized in Table 17.

Table 16. **Summary of findings for helping situations**

Hypothesis	Level of Support	
	Scoring method 1	Scoring method 2
Hypothesis 1: Respondent race will moderate the effects of the race mismatch between the respondent and the non-focal actor on helping, such that the effects of race mismatch will be more negative for White respondents viewing a minority non-focal actor than for minority respondents viewing a White non-focal actor.	No support	No support
Hypothesis 2: There will be a main effect for respondent sex on helping, such that female respondents will offer more help to non-focal actors than will male respondents.	No support	Partial support (HS1 options 3 and 6 (marginal); HS2 options 5 and 2 (marginal))
Hypothesis 3: Respondent sex will moderate the effects of non-focal actor sex on helping, such that a female non-focal actor will receive more help than a male non-focal actor when the respondent is male. Female and male non-focal actors will receive equal amounts of help when the respondent is female.	Supported	No support
Hypothesis 4: In a helping situation where the competence stereotype is activated, the lowest level of help will be offered by White male respondents to male Black non-focal actors as compared to other race and sex-based combinations of respondents with targets.	Could not test	No support
Hypothesis 5a: In a helping situation where the work ethic stereotype is activated, the lowest level of help will be offered by White respondents to Black non-focal actors as compared to other race-based combinations of respondents with targets.	No support	No support
Hypothesis 5b: In a helping situation where the work ethic stereotype is activated, the highest level of help will be offered to Asian non-focal actors.	No support	No support

Note. HS1 = Helping situation 1. HS2 = Helping situation 2.

Table 17. **Effects of demographic cues on performance on helping items**

Test	Finding	
	DV: Helping situation 1 score	DV: Helping situation 2 score
Test 1: Does respondent race moderate the effects of the race mismatch between the respondent and the non-focal actor on scores?	No; marginal negative effect of race mismatch on both minority and White respondents	Yes; more negative effects of race mismatch on minority than on White respondents
Test 2: Is there a main effect for respondent sex on scores?	No	No
Test 3: Does respondent sex moderate the effects of non-focal actor sex on scores?	No	No
Test 4: Is there a four-way interaction of respondent sex, dichotomized respondent race (minority, White), non-focal actor sex, and non-focal actor race (Black, White) on helping situation 1 score?	No; three-way interaction of respondent sex, non-focal actor sex, and non-focal actor race was significant	--
Test 5a: Is there an interaction of respondent race (Asian, Black, White) and non-focal actor race (Asian, Black, White) on helping situation 2 score?	--	Yes; Black respondents scored lower when the non-focal actor was Asian as opposed to Black
Test 5b: Does non-focal actor race (Asian, Black, White) affect helping situation 2 score?	--	No

Note. Test numbers correspond roughly to hypothesis tests in Table 16.

Tests of Hypotheses Pertaining to Conflict Situations

Hypotheses 6a through 16 could not be tested using scoring method 1 due to the lack of internal consistency of response options pertaining to both of the conflict situations. Before discussing hypotheses pertaining to conflict, I include Tables 18 and 19 below to provide an overview of mean likelihood ratings as a function of sex (Table 18) and race (Table 19) of the respondent and non-focal actor.

Table 18. **Likelihood ratings for conflict response options as a function of respondent's and non-focal actor's sex**

	Female Respondent				Male Respondent			
	Female non-focal actor (*N* = 96)		Male non-focal actor (*N* = 93)		Female non-focal actor (*N* = 95)		Male non-focal actor (*N* = 90)	
Response Option	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CS1 Op1: Tell the coworker that you will continue to watch for unglued cartons and get back to him/her.	4.48	.74	4.14	.89	4.43	.61	4.09	.92
CS1 Op2: Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.	2.55	1.02	2.46	1.06	2.44	1.08	2.60	1.06
CS1 Op3: Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.	3.36	1.09	3.67	1.00	3.47	1.06	3.42	1.04
CS1 Op4: Ask another team member to discuss the problem with your coworker.	2.81	1.17	2.66	1.17	2.87	1.04	2.74	1.13
CS1 Op5: Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.	3.15	1.26	3.03	1.17	3.06	1.15	2.87	1.19
CS2 Op1: Ask your coworker what you could do to get along better.	4.15	.98	3.87	1.19	3.71	1.20	3.97	.97
CS2 Op2: Ask the team leader for suggestions about how to work with your coworker.	3.95	.99	3.76	1.07	3.64	1.12	3.51	1.11
CS2 Op3: Try to avoid your coworker whenever possible because you can't change him/her.	2.36	1.05	2.60	1.15	2.59	1.09	2.34	1.17
CS2 Op4: Tell your coworker that he/she should act like a team member and try to work together.	3.52	1.10	3.53	1.10	3.47	1.14	3.56	1.03
CS2 Op5: Ask the team leader to meet with you and your coworker to work out your differences.	3.48	1.26	3.19	1.33	3.29	1.30	3.10	1.25
CS2 Op6: Ask other team members for suggestions on how to work with your coworker.	3.58	1.18	3.70	1.17	3.67	1.03	3.82	1.01

Note. CS1 = Conflict situation 1. CS2 = Conflict situation 2.

Table 19. **Likelihood ratings for conflict response options as a function of respondent's and non-focal actor's race**

	Asian Respondent						Black Respondent						White Respondent					
	Asian non-focal actor (*N* = 17)		Black non-focal actor (*N* = 12)		White non-focal actor (*N* = 12)		Asian non-focal actor (*N* = 11)		Black non-focal actor (*N* = 14)		White non-focal actor (*N* = 10)		Asian non-focal actor (*N* = 94)		Black non-focal actor (*N* = 89)		White non-focal actor (*N* = 88)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CS1.1	4.24	.66	3.92	1.31	3.75	1.22	4.55	.52	4.64	.50	4.20	.63	4.34	.84	4.30	.74	4.32	.75
CS1.2	2.76	1.15	2.17	1.27	3.25	1.29	2.36	1.12	2.36	1.22	2.90	1.10	2.49	1.03	2.34	.90	2.65	1.08
CS1.3	3.82	.73	3.50	1.24	3.67	.89	3.27	.90	4.07	.73	4.10	.74	3.45	1.03	3.43	1.02	3.19	1.15
CS1.4	3.12	1.17	3.00	1.28	3.17	1.34	3.00	.89	2.64	1.22	2.80	1.48	2.70	1.08	2.82	1.13	2.63	1.11
CS1.5	3.00	1.37	2.92	1.51	2.92	1.31	3.45	1.13	3.14	1.23	3.10	1.52	3.02	1.22	3.18	1.08	2.76	1.15
CS2.1	3.88	1.22	3.50	1.17	4.08	1.08	3.55	1.44	3.93	.92	3.70	1.25	3.89	1.18	4.12	.91	3.88	1.12
CS2.2	3.47	1.23	3.83	1.19	3.50	1.57	3.91	.83	3.79	1.12	3.60	1.17	3.74	1.07	3.87	1.08	3.67	.97
CS2.3	2.24	.75	2.92	1.44	2.75	1.22	2.73	1.19	1.71	.83	2.20	1.40	2.38	1.09	2.48	1.06	2.57	1.08
CS2.4	3.71	.69	3.08	1.38	3.75	.87	3.73	1.42	3.29	1.27	4.40	.52	3.38	1.13	3.56	1.09	3.42	1.08
CS2.5	3.29	1.10	2.50	1.45	3.50	1.09	4.09	1.22	3.64	1.39	3.00	1.33	3.29	1.23	3.38	1.23	2.97	1.37
CS2.6	3.94	.43	4.00	1.04	4.17	.39	3.64	1.21	3.00	1.41	3.00	1.63	3.72	1.08	3.61	1.16	3.75	1.02

Note. CS1 = Conflict situation 1. CS2 = Conflict situation 2.

H6, scoring method 2. Method 2 was used to evaluate Hypothesis 6, that in responses to conflict situations there will be a main effect for respondent-non-focal actor race match on assertiveness (6a) and cooperativeness (6b), in an exploratory manner (by using unweighted likelihood ratings). A MANOVA was performed to examine the impact of race match on likelihood ratings for the 11 response options associated with the 2 conflict items. There were no significant differences in likelihood ratings based on whether the respondent's and non-focal actor's race matched or not, Wilks' lambda = .97, $F(11,285) = .74$, ns. This likely indicates that Hypothesis 6 would not have received support if the more appropriate scoring method 1 (with weighted scores) was used.

H6, scoring method 3. Method 3 was then applied to test for main effects of respondent-non-focal actor race match on effectiveness (as defined by the test developer) of responses to conflict situations 1 and 2. There was no significant main effect for race match on neither conflict item 1 score nor conflict item 2 score, $F(1,318) = .39$, $MSE = 1.01$, ns, and $F(1,316) = .18$, $MSE = .70$, ns, respectively.

H7, scoring method 2. To evaluate Hypothesis 7, that there will be an interaction between respondent race and respondent-non-focal actor race match on preference for third party intervention when the goal is reconciliation, in an exploratory manner, a MANOVA was again used. The 6 response options associated with conflict situation 2 (where the goal is reconciliation) were entered in as the dependent variables. The analysis showed no statistically significant interaction of respondent race (minority, White) and race match of respondent and non-focal actor on likelihood ratings, Wilks' lambda = .97, $F(6,309) = 1.82$, $p = .096$. Thus, Hypothesis 7 was not supported using scoring method 2.

H7, scoring method 3. Scoring method 3 was next applied to test for an interaction of respondent race and respondent-non-focal actor race match on effectiveness of responses to conflict situation 2. The ANOVA did not show a statistically significant interaction, $F(1,314) = 1.02$, $MSE = .70$, ns.

H8, scoring method 2. Next, scoring method 2 was applied to assess Hypothesis 8, that there will be an interaction between respondent race and respondent-non-focal actor race match on preference for third party intervention when the goal is reporting a coworker's poor performance. The MANOVA with the 5 response options associated with conflict situation 1 (where the goal of intervention is reporting a coworker's poor performance) entered in as dependent variables showed no statistically significant interaction of respondent race (minority, White) and race match of respondent and non-focal actor on likelihood ratings, Wilks' lambda = .97, $F(5,312) = 1.72$, ns. It is therefore likely that Hypothesis 8 would not have been supported if it were possible to test it directly (using method 1) as had been planned.

H8, scoring method 3. Scoring method 3 was applied to test for an interaction of respondent race and respondent-non-focal actor race match on effectiveness of responses to conflict situation 1. The ANOVA did not show a statistically significant interaction, $F(1,316) = 1.13$, $MSE = 1.01$, ns.

H9, scoring method 2. Next, Hypothesis 9, that female respondents will respond to conflict with lower assertiveness (9a) and higher cooperativeness (9b) than will male respondents, was explored. A MANOVA was performed to examine the impact of respondent gender on likelihood ratings for the 11 response options associated with the 2 conflict items. There were no significant differences in likelihood ratings based on respondent gender, Wilks'

lambda = .97, $F(11,309) = .77$, ns. This may indicate that Hypothesis 9 would not have received support if it were possible to test it using scoring method 1 as planned.

H9, scoring method 3. Next, the main effect of gender on effectiveness of responses to conflict situations 1 and 2 was evaluated using scoring method 3. There was no significant main effect of gender on either item, $F(1,343) = .48$, $MSE = 1.00$, ns, and $F(1,344) = .03$, $MSE = .72$, ns, respectively.

H10, scoring method 2. Hypothesis 10, that respondent gender will moderate the effects of gender match between the respondent and the non-focal actor on assertiveness (10a) and cooperativeness (10b) when responding to conflict was addressed next. A MANOVA showed a statistically significant effect of the gender by gender match interaction on likelihood ratings for the 11 response options associated with the 2 conflict items, Wilks' lambda = .93, $F(11,307) = 2.05$, $p < .05$. ANOVAs were used to further examine likelihood ratings for the two response options that appeared to be influenced by the gender by gender match interaction. An ANOVA on option 1 of conflict situation 1 (Tell your coworker you will continue to watch for unglued cartons and get back to him/her) showed a statistically significant effect of the interaction on likelihood ratings, $F(1, 341) = 13.55$, $MSE = .63$, $p < .001$. Separate ANOVAs by gender of respondent showed that men endorse response option 1 significantly less strongly when the coworker is of the same gender (i.e., male), $F(1,168) = 7.84$, $MSE = .59$, $p < .01$, while women endorse response option 1 significantly *more* strongly when the coworker is of the same gender (i.e., female), $F(1,173) = 5.85$, $MSE = .67$, $p < .05$.

Given that a higher willingness to respond to the unglued cartons problem by telling the coworker that he/she will continue to monitor the cartons situation (and not really solve the problem) corresponds to a low level of assertiveness according to judges' ratings ($M = 1.88$ on a

3-pt scale, *SD* = .91; lower mean than for 3 of the 4 other response options), this result is consistent with the expectation that gender match will have a positive effect on assertiveness for male respondents and a negative effect on assertiveness for female respondents. On the flipside, because a higher willingness to respond to the unglued cartons problem this way corresponds to a higher level of cooperativeness according to judges' ratings (*M* = 2.55 on a 3-pt scale, *SD* = .78), this result is consistent with the expectation that gender match will have a negative effect on cooperativeness for male respondents and a positive effect on cooperativeness for female respondents. Thus, both male and female respondents are more willing to not be very assertive (or to be more cooperative) with a coworker over a cartons situation when the coworker is female. These results with respect to conflict situation 1's response option 1, then, lend some amount of support to Hypothesis 10.

Next, an ANOVA on option 4 of conflict situation 1 (Ask another team member to discuss the problem with your coworker) showed the effect of the interaction of gender and gender match failing to reach a statistically significant effect on likelihood ratings, $F(1,341) = 2.53$, $MSE = 1.25$, $p = .113$. A closer look at the cell means, however, revealed that they trended toward the same direction as those for response option 1 discussed above: respondents were somewhat more likely to endorse response option 4 when the coworker was female as opposed to male. Yet, the results with respect to conflict situation 1's response option 4 do not lend support to Hypothesis 10 in a statistically significant manner.

H10, scoring method 3. Next, scoring method 3 was applied to examine whether respondent gender moderated the effects of gender match between the respondent and the non-focal actor on effectiveness of responses to the conflict items. The interaction of respondent gender with gender match on conflict item 1 score was not statistically significant, $F(1,341) =$

.81, *MSE* = 1.00, ns. The effect on conflict item 2 score was likewise not significant, $F(1,342) = 2.28$, *MSE* = .72, ns.

H11, scoring method 2. To explore the tenability of Hypothesis 11, that when the goal of the conflict situation is reconciliation both male and female respondents will prefer third party intervention more when the non-focal actor is female, a MANOVA with the 6 response options associated with conflict situation 2 (where the goal is reconciliation) entered in as dependent variables was calculated. The gender by gender match interaction did not have a significant effect on likelihood ratings on the 6 response options associated with conflict situation 2, Wilks' lambda = .98, $F(6,337) = .90$, ns. This may indicate that Hypothesis 11 would not have received support if it were possible to test it as planned. And as already reviewed above, there is also no significant interaction of respondent gender with gender match on effectiveness of responses to conflict situation 2.

H12, scoring method 2. A MANOVA was likewise used to explore Hypothesis 12, that when the goal of the conflict situation is reporting a coworker's poor performance, both male and female respondents will show less preference for third party intervention when the non-focal actor is female. There was a statistically significant effect of the gender by gender match interaction on likelihood ratings for the 5 response options associated with conflict situation 1 (where the goal of intervention would be to report a coworker's poor performance), Wilks' lambda = .94, $F(5,337) = 4.07$, $p = .001$. Next, the follow up analysis for option 1 of conflict situation 1 (Tell your coworker you will continue to watch for unglued cartons and get back to him/her) matches that of Hypothesis 10 above and the results are exactly the same: men endorse response option 1 significantly less strongly when the coworker is of the same gender (i.e., male), $F(1,168) = 7.84$, *MSE* = .59, $p < .01$, while women endorse response option 1

significantly *more* strongly when the coworker is of the same gender (i.e., female), $F(1,173) = 5.85$, $MSE = .67$, $p < .05$. In the context of the analysis of Hypothesis 12, this suggests that both male and female respondents are less willing to get a third party to intervene in an unglued cartons situation when the coworker is female. This result for response option 1 of conflict situation 1 is consistent with what Hypothesis 12 predicts will happen. As already reviewed earlier, there is no significant interaction of respondent gender with gender match on effectiveness of responses to conflict situation 1, so method 3 was not applied to test again for an interaction on conflict item 1 scores.

H13, scoring method 2. Hypothesis 13 predicted that in conflict situations the lowest level of assertiveness will be exhibited by female respondents paired with a female non-focal actor of a different race (13a) and that the highest level of assertiveness will be exhibited by male respondents paired with a male non-focal actor of the same race (13b), as compared to other sex and race-based combinations of respondents with targets. To explore these proposed relationships, gender of respondent, gender of non-focal actor, and race match of respondent relative to the non-focal actor (match, mismatch) were entered as independent variables into a MANOVA. The analysis did not show a statistically significant effect of the three way interaction on the likelihood ratings respondents gave to the 11 response options associated with the two conflict situations, Wilks' lambda = .96, $F(11,279) = .97$, ns. The main effect for non-focal actor gender was significant, Wilks' lambda = .93, $F(11,279) = 2.04$, $p < .05$. This was due to conflict item 1's response option 1 being rated differently by respondents depending on whether the non-focal actor was male or female. This finding was discussed earlier in the context of Hypothesis 10. Thus, because the three way interaction was not significant, there was no

support for Hypotheses 13a or 13b based on an analysis using likelihood ratings for individual response options.

H14, scoring method 2. Hypothesis 14 proposed that in conflict situations the lowest level of cooperativeness will be exhibited by male respondents paired with male non-focal actors of a different race (14a) and that the highest level of cooperativeness will be exhibited by female respondents paired with a female non-focal actor of the same race (14b) as compared to other sex and race-based combinations of respondents with targets. An exploration of this hypothesis at the individual response option level looks similar to that of Hypothesis 13 above. Thus, there was no support for Hypotheses 14a and 14b based on an analysis using likelihood ratings for individual response options.

H14, scoring method 3. Next, method 3 was applied to examine whether the interaction of gender of respondent, gender of non-focal actor, and race match of respondent relative to the non-focal actor (match, mismatch) affected effectiveness of responses to conflict situations 1 and 2. The 3-way interaction on conflict item 1 score was not significant, $F(1,312) = 2.18$, $MSE = 1.01$, ns, and neither was the 3-way interaction on conflict item 2 score, $F(1,310) = .04$, $MSE = .70$, ns. None of the lower order interactions or main effects were statistically significant either.

H15, scoring method 2. Hypothesis 15 proposed that in a conflict situation where the goal of involving a third party is reconciliation, the highest preference for third party intervention will be exhibited by collectivist (Asian, Black) respondents paired with a female non-focal actor of the same race. To explore this proposed relationship a MANOVA was calculated. The independent variables entered into the analysis were dichotomized respondent race (minority, White), non-focal actor sex, and race match between respondent and non-focal actor (match, mismatch). The proposed three way interaction on the six response options associated with

conflict item 2 (where the goal of third party intervention would be reconciliation) was not statistically significant, Wilks' lambda = .99, $F(6,305) = .46$, ns. None of the lower order interactions or main effects were significant either. These results at the level of the response options do not lend support to Hypothesis 15.

H15, scoring method 3. Next, scoring method 3 was used to check for a three way interaction of dichotomized respondent race (minority, White), non-focal actor sex, and race match between respondent and non-focal actor (match, mismatch) on effectiveness of responses to conflict item 2. The three way interaction on conflict item 2 score was marginally significant, $F(1,310) = 3.66$, $MSE = .69$, $p = .057$. A lower order interaction of respondent race with non-focal actor sex was also significant, $F(1,310) = 6.74$, $MSE = .69$, $p = .01$. The main effect for non-focal actor sex was significant as well, $F(1,310) = 5.13$, $MSE = .69$, $p < .05$. To break down the significant two-way interaction separate ANOVAs were calculated by dichotomized respondent race. White respondents scored the same on the item regardless of whether the non-focal actor was female ($M = .57$, $SD = .84$) or male ($M = .55$, $SD = .81$). Minority (Asian, Black) respondents who saw a female non-focal actor scored higher on conflict item 2 ($M = .74$, $SD = .76$) than minority respondents who saw a male non-focal actor ($M = .26$, $SD = .92$), $F(1,71) = 5.95$, $MSE = .71$, $p < .05$. To try to understand this effect of non-focal actor gender on minority respondents' conflict situation 2 scores, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses. Non-focal actor gender did not have a statistically significant effect on the frequency with which various response options were selected as most likely, $\chi^2(5, N = 73) = 8.76$, ns, or least likely, $\chi^2(5, N = 73) = 9.74$, ns. Therefore, actual frequencies of responses were not examined.

H16, scoring method 2. Next, Hypothesis 16 was explored using scoring method 2. It stated that in a conflict situation where the goal of involving a third party is reporting a coworker's poor performance, the highest preference for third party intervention will be exhibited by collectivist (Asian, Black) respondents paired with a male non-focal actor of a different race. A MANOVA was calculated using the response options associated with conflict situation 1 (where the goal of third party intervention would be reporting a coworker's poor performance). The independent variables entered into the analysis were dichotomized respondent race (minority, White), non-focal actor sex, and race match between respondent and non-focal actor (match, mismatch). The proposed three way interaction on the five response options associated with conflict item 1 was not statistically significant, Wilks' lambda = .99, $F(5,308)$ = .604, ns. The main effects for dichotomized respondent race and for non-focal actor sex were significant, Wilks' lambda = .96, $F(5,308) = 2.66$, $p < .05$, and Wilks' lambda = .96, $F(5,308)$ = 2.45, $p < .05$, respectively. As mentioned several times before, the main effect for non-focal actor sex is attributable to conflict item 1's response option 1 being rated differently by respondents depending on whether the non-focal actor was male or female. The main effect for dichotomized respondent race had to do with minorities giving higher likelihood ratings on average ($M = 3.79$, $SD = .85$) to response option 3 (Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse) than White respondents ($M = 3.39$, $SD = 1.07$), $F(1,321) = 8.34$, $MSE = 1.06$, $p < .01$. These results at the level of the response options do not lend support to Hypothesis 16.

H16, scoring method 3. Scoring method 3 was used to check for a three way interaction of dichotomized respondent race (minority, White), non-focal actor sex, and race match between respondent and non-focal actor (match, mismatch) on effectiveness of responses to conflict item

1. The three way interaction on conflict item 1 score was not statistically significant, $F(1,312) = .07$, $MSE = .99$, ns. The main effect for non-focal actor sex was significant, $F(1,312) = 4.08$, $MSE = .99$, $p < .05$, and so was the two-way interaction of dichotomized respondent race and non-focal actor sex, $F(1,312) = 6.85$, $MSE = .99$, $p < .01$. The nature of this two-way interaction was such that White respondents' scores were not impacted by whether the non-focal actor was female ($M = 1.19$, $SD = 1.02$) or male ($M = 1.28$, $SD = .95$), $F(1,249) = .54$, $MSE = .97$, ns, whereas minority (Asian, Black) respondents scored higher on conflict item 1 when the non-focal actor was female ($M = 1.42$, $SD = .83$) than when that actor was male ($M = .76$, $SD = 1.18$), $F(1,70) = 7.58$, $MSE = 1.02$, $p < .01$. These results are consistent with those reported above (H15, scoring method 3) for conflict item 2. To try to understand this effect of non-focal actor gender on minority respondents' conflict situation 1 scores, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses. Non-focal actor gender did not have a statistically significant effect on the frequency with which various response options were selected as most likely, $\chi^2(4, N = 72) = 3.70$, ns, and the effect on least likely responses was only marginally significant, $\chi^2(4, N = 72) = 9.37$, $p = .053$. Therefore, actual frequencies of responses were not examined.

Next, results pertaining to Hypotheses 17 and 18 (about applicant reactions) are reviewed. Before proceeding, results for Hypotheses 6 through 16 (about conflict) are summarized in Table 20 below. Also, the results of the analyses pertaining to response *effectiveness* are summarized in Table 21.

Table 20. **Summary of findings for conflict situations**

Hypothesis	Level of Support[a]
Hypothesis 6a: In responses to conflict situations respondents will endorse a more assertive response when the respondent and the non-focal actor are of the same (vs different) race.	No support
Hypothesis 6b: In responses to conflict situations respondents will endorse a more cooperative response when the respondent and the non-focal actor are of the same (vs different) race.	No support
Hypothesis 7: When the goal is reconciliation respondents belonging to collectivist groups (Asians, Blacks) will show a preference for third party intervention when paired with non-focal actors of the same (vs different) race as them. Race match between respondents and non-focal actors will not influence preferences for third party intervention among individualist respondents (Whites).	No support
Hypothesis 8: When the goal of third party intervention is reporting a coworker's poor performance respondents belonging to collectivist groups (Asians, Blacks) will show a preference for third party intervention when paired with non-focal actors of a different (vs same) race as them. Race match between respondents and non-focal actors will not influence preferences for third party intervention among individualist respondents (Whites).	No support
Hypothesis 9a: Female respondents will respond to conflict with lower assertiveness than male respondents.	No support
Hypothesis 9b: Female respondents will respond to conflict with higher cooperativeness than male respondents.	No support
Hypothesis 10a: For male respondents, gender match with the non-focal actor will have a positive effect on assertiveness. For female respondents, gender match will have a negative effect on assertiveness.	Minimal (CS1 option 4)
Hypothesis 10b: For male respondents, gender match with the non-focal actor will have a negative effect on cooperativeness. For female respondents, gender match will have a positive effect on cooperativeness.	Minimal (CS1 option 4)
Hypothesis 11: When the goal of the conflict situation is reconciliation: a. Female respondents will show a preference for third party intervention when paired with female (vs male) non-focal actors; b. Male respondents will show a preference for third party intervention when paired with female (vs male) non-focal actors.	No support
Hypothesis 12: When the goal of the conflict situation is reporting a coworker's poor performance: a. Female respondents will show *less* of a preference for third party intervention when paired with female (vs male) non-focal actors; b. Male respondents will show *less* of a preference for third party intervention when paired with female (vs male) non-focal actors.	Minimal (CS1 option 1)

Table 20 (cont'd)

Hypothesis 13a: In conflict situations the lowest level of assertiveness will be exhibited by female respondents paired with a female non-focal actor of a different race as compared to other sex and race-based combinations of respondents with targets.	No support
Hypothesis 13b: In conflict situations the highest level of assertiveness will be exhibited by male respondents paired with a male non-focal actor of the same race as compared to other sex and race-based combinations of respondents with targets.	No support
Hypothesis 14a: In conflict situations the lowest level of cooperativeness will be exhibited by male respondents paired with male non-focal actors of a different race as compared to other sex and race-based combinations of respondents with targets.	No support
Hypothesis 14b: In conflict situations the highest level of cooperativeness will be exhibited by female respondents paired with a female non-focal actor of the same race as compared to other sex and race-based combinations of respondents with targets.	No support
Hypothesis 15: In a conflict situation where the goal of involving a third party is reconciliation, the highest preference for third party intervention will be exhibited by collectivist (Asian, Black) respondents paired with a female non-focal actor of the same race.	No support
Hypothesis 16: In a conflict situation where the goal of involving a third party is reporting a coworker's poor performance, the highest preference for third party intervention will be exhibited by collectivist (Asian, Black) respondents paired with a male non-focal actor of a different race.	No support

Note .[a]Hypotheses were tested using scoring method 2; scoring method 1 could not be applied. CS1 = Conflict situation 1.

Table 21. **Effects of demographic cues on performance on conflict items**

Test	Finding	
	DV: Conflict situation 1 score	DV: Conflict situation 2 score
Test 6: Does race match between the respondent and the non-focal actor affect scores?	No	No
Test 7/8: Is there an interaction of respondent race and race match between respondent and non-focal actor on scores?	No	No
Test 9: Is there a main effect for respondent gender on scores?	No	No
Test 10/11/12: Is there an interaction of respondent gender and gender match between respondent and non-focal actor on scores?	No	No
Test 13/14: Is there a three-way interaction of respondent gender, non-focal actor gender, and race match between respondent and non-focal actor on scores?	No	No
Test 15/16: Is there a three way interaction of dichotomized respondent race (minority, White), non-focal actor gender, and race match between respondent and non-focal actor on scores?	No; two-way interaction of dichotomized respondent race and non-focal actor gender was significant	No; two-way interaction of dichotomized respondent race and non-focal actor gender was significant

Note. Test numbers correspond roughly to hypothesis tests in Table 20.

Correlations with Applicant Reactions

Looking at the correlations between dichotomized respondent race and applicant reactions shown in Table 10, minorities apparently perceived higher opportunity to perform on the test (r = -.13) and higher job relatedness of the test (both predictive and content-based; r = -.14 and r = -.13, respectively). T-tests (on the respondents who passed all four manipulation checks) confirmed that as a group, Asian and Black respondents reported higher opportunity to perform (M = 3.18, SD = .86) than White respondents (M = 2.85, SD = .95), $t(276) = 2.48$, $p < .05$, as well as higher job relatedness-predictive (M = 2.31 and SD = 1.09 compared to M = 1.91 and SD = .84), $t(86.65) = 2.74$, $p < .01$, and higher job relatedness-content (M = 2.36 and SD = .95 compared to M = 2.01 and SD = .88), $t(276) = 2.74$, $p < .01$. However, when ANOVAs were used to examine differences by respondent race (Asian, Black, White), post hoc Tukey tests showed none of the pairwise differences to be statistically significant.

Looking again at the correlations in Table 10, respondent gender was correlated with perceived compatibility with the organization (r = .17), indicating more positive reactions from male respondents. T-tests (on the respondents who passed all four manipulation checks) likewise showed that male respondents reported higher compatibility with the organization (M = 3.32, SD = .95) than female respondents (M = 2.93, SD = .84), $t(296) = -3.80$, $p < .001$. Additionally, the difference between men and women on attraction toward the organization was marginally significant, $t(296) = -1.96$, $p = .051$, with men again reacting more positively (M = 3.71, SD = .85) than women (M = 3.53, SD = .78).

Tests of Hypotheses Pertaining to Applicant Reactions

Before discussing hypotheses pertaining to applicant reactions, I include Tables 22 and 23 below to provide an overview of reactions as a function of sex (Table 22) and race (Table 23) of the respondent and non-focal actor.

Table 22. **Reactions as a function of respondent's and non-focal actor's sex**

	Female Respondent						Male Respondent					
	Female non-focal actor (N = 96)		Male non-focal actor (N = 93)		Total (N = 189)		Female non-focal actor (N = 95)		Male non-focal actor (N = 90)		Total (N = 185)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Opportunity to Perform	2.94	.98	2.94	.87	2.94	.92	2.93	.98	2.96	.92	2.95	.95
Job Relatedness-Predictive	2.05	1.05	1.99	.83	2.02	.94	2.01	.89	1.95	.89	1.98	.89
Job Relatedness-Content	2.06	.98	2.09	.94	2.07	.96	2.23	.97	2.22	.95	2.22	.96
Attraction	3.59	.77	3.49	.81	3.54	.79	3.67	.92	3.71	.80	3.69	.86
Compatibility	3.01	.82	2.99	.84	3.00	.83	3.27	.99	3.33	.90	3.30	.95

Table 23. **Reactions as a function of respondent's and non-focal actor's race**

	Asian Respondent						Black Respondent						White Respondent					
	Asian non-focal actor (*N* = 17)		Black non-focal actor (*N* = 12)		White non-focal actor (*N* = 12)		Asian non-focal actor (*N* = 11)		Black non-focal actor (*N* = 14)		White non-focal actor (*N* = 10)		Asian non-focal actor (*N* = 94)		Black non-focal actor (*N* = 89)		White non-focal actor (*N* = 88)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Opportunity to Perform	3.09	.78	3.42	.62	3.06	.91	3.18	.73	3.46	.81	2.80	1.10	2.82	.96	2.99	.98	2.84	.93
Job Related.-Predictive	2.26	.97	2.13	1.07	2.25	.94	2.82	1.12	1.96	1.03	1.95	1.04	2.02	.96	1.93	.81	1.86	.84
Job Related.-Content	2.47	.86	1.83	.98	2.38	.77	2.23	.88	2.61	1.15	2.25	.82	2.01	.91	2.18	.92	1.95	.89
Attraction	3.66	.68	3.07	.60	3.79	.69	3.62	.84	3.92	.79	3.52	.45	3.58	.87	3.61	.80	3.66	.86
Compatibility	3.00	.91	2.77	.64	3.00	.76	3.11	.72	3.43	1.06	2.95	.62	3.16	.93	3.15	.87	3.18	.88

H17. Hypothesis 17 stated that minority respondents will react more positively when the SJT features Black or Asian non-focal actors than when the SJT features White non-focal actors. Because the dependent variables, attraction toward the organization and perceived compatibility with the organization, were both significantly correlated with performance on the test (r = .23 and r = .18, respectively; see Table 10), total SJT score was controlled for in the analyses. For consistency, SJT score was likewise controlled for in the analyses with opportunity to perform and job relatedness of the test. Further, the analysis excluded individuals who failed one or more of the manipulation checks (i.e., identifying the race and gender of the actors). An ANCOVA showed a statistically significant effect of non-focal actor race on minority respondents' perceived opportunity to perform, $F(1,61) = 4.51$, MSE = .71, $p < .05$. Minority respondents perceived better opportunity to perform when the non-focal actor was a minority (Asian, Black) (M = 3.31, SD = .75) as opposed to White (M = 2.79, SD = 1.04). ANCOVAs showed no statistically significant effects of non-focal actor race on minority respondents' perceived predictive-type job relatedness of the test, $F(1,61) = .27$, MSE = 1.10, ns, content-type job relatedness of the test, $F(1,61) = .07$, MSE = .93, ns, attraction toward the organization, $F(1,61)$ = .03, MSE = .53, ns, and on perceived compatibility with the organization, $F(1,61) = .27$, MSE = .86, ns. Hypothesis 17 was partially supported.

It should also be noted that ANCOVAs showed no statistically significant effects of non-focal actor race on *White* respondents' perceived opportunity to perform, $F(1,211) = .02$, MSE = .87, ns, predictive-type job relatedness of the test, $F(1,211) = .80$, MSE = .71, ns, content-type job relatedness of the test, $F(1,211) = .35$, MSE = .77, ns, attraction toward the organization, $F(1,211) = 1.53$, MSE = .66, ns, and on perceived compatibility with the organization, $F(1,211)$ = .65, MSE = .71, ns.

H18. Hypothesis 18 predicted that female respondents will react more positively when the non-focal actors are male. ANCOVAs showed no statistically significant effects of non-focal actor gender on women's perceived opportunity to perform, $F(1,143) = .36$, $MSE = .83$, ns, predictive-type job relatedness of the test, $F(1,143) = .85$, $MSE = .92$, ns, content-type job relatedness of the test, $F(1,143) = .46$, $MSE = .86$, ns, attraction toward the organization, $F(1,143) = 2.77$, $MSE = .60$, ns, and on perceived compatibility with the organization, $F(1,143) = .84$, $MSE = .70$, ns. Hypothesis 18 was not supported.

It should also be noted that ANCOVAs likewise showed no statistically significant effects of non-focal actor gender on men's perceived opportunity to perform, $F(1,149) = 3.49$, $MSE = .87$, $p = .064$, predictive-type job relatedness of the test, $F(1,149) = .87$, $MSE = .78$, ns, content-type job relatedness of the test, $F(1,149) = .33$, $MSE = .92$, ns, attraction toward the organization, $F(1,149) = .00$, $MSE = .67$, ns, and on perceived compatibility with the organization, $F(1,149) = .07$, $MSE = .85$, ns.

Considering the overall findings with regard to applicant reactions, as a group, minorities perceived better opportunity to perform when the non-focal actor was a minority (Asian, Black), and relative to White respondents, they perceived higher job relatedness (content and predictive) of the test in general. White respondents did not react to the test differently based on the race of non-focal actors. Female respondents did not react to the test differently based on the gender of non-focal actors but perceived worse compatibility with the organization in general relative to male respondents. Female respondents did not differ significantly from male respondents on any of the other four reactions measures. Before discussing exploratory analyses using overall test score, results for hypotheses pertaining to applicant reactions are summarized in Table 24 below.

Table 24. **Summary of findings for applicant reactions**

Hypothesis	Level of Support
Hypothesis 17: Minority respondents (Asian, Black) will react more positively when the SJT features Black or Asian non-focal actors than when the SJT features White non-focal actors.	Partial support (minority respondents perceived better opportunity to perform when non-focal actor was a minority)
Hypothesis 18: Female respondents will react more positively when the non-focal actors are male.	No support

Exploratory Analyses Using Overall Test Score

As described earlier, respondents' most and least likely responses to the four SJT items were scored according to the scoring program originally used by the organization that provided their SJT for this study. Respondents' scores ranged from -3 to 8.

The correlations of test score with other variables in the study are presented in Table 10. SJT scores are positively associated with perceptions of opportunity to perform (r = .15), attraction to the hiring organization (r = .23), and perceived compatibility with the hiring organization (r = .18). SJT scores are not related to perceptions of the job relatedness of the test though.

As a first step to exploring the variables that impact overall SJT scores, a four way interaction of the demographic (respondent race, respondent gender) and relational demographic (race match between respondent and non-focal actor, gender match between respondent and non-focal actor) variables in this study was tested. An ANOVA showed the four way interaction of these variables on SJT score to not be statistically significant, $F(2,250) = .75$, $MSE = 3.53$, ns. However, a lower order interaction of respondent race by race match between respondent and non-focal actor, as well as a main effect for race match were statistically significant ($p < .05$). The ANOVA results are displayed in Table 25.

Table 25. **Results of ANOVA for effects of race, gender, race match, gender match, and their interactions on test scores**

	SJT Score		
Effect	*F*	*p*	Partial η^2
Race (R): Asian, Black, White	.23	.79	.00
Gender (G): female, male	.01	.90	.00
Race match (RM): different race, same race	6.24	.01	.02
Gender match (GM): different gender, same gender	1.64	.20	.01
R x G	.18	.84	.00
R x RM	4.01	.02	.03
R x GM	2.07	.13	.02
G x RM	.00	.99	.00
G x GM	.38	.54	.00
RM x GM	.68	.41	.00
R x G x RM	.11	.90	.00
R x G x GM	1.54	.22	.01
R x RM x GM	.89	.41	.01
G x RM x GM	.12	.72	.00
R x G x RM x GM	.75	.47	.01

Note. Total $n = 274$. Respondents who were not uniquely Asian, Black, or White ($n = 27$) and respondents who failed one or more manipulation check ($n = 76$) were excluded.

To break down the race by race match interaction, separate ANOVAs by race of respondent were calculated. As can be seen in Figure 22, there was no statistically significant effect of race match on Asian respondents, $F(1,34) = .11$, $MSE = 4.24$, ns, and no statistically significant effect of race match on White respondents, $F(1,221) = .05$, $MSE = 3.32$, ns. There was, however, a statistically significant effect of race match on Black respondents, $F(1,32) = 6.78$, $MSE = 3.40$, $p < .05$. Black respondents scored higher on the SJT when the non-focal actor was Black ($M = 3.76$, $SD = 1.79$) as opposed to White or Asian ($M = 2.12$, $SD = 1.90$).

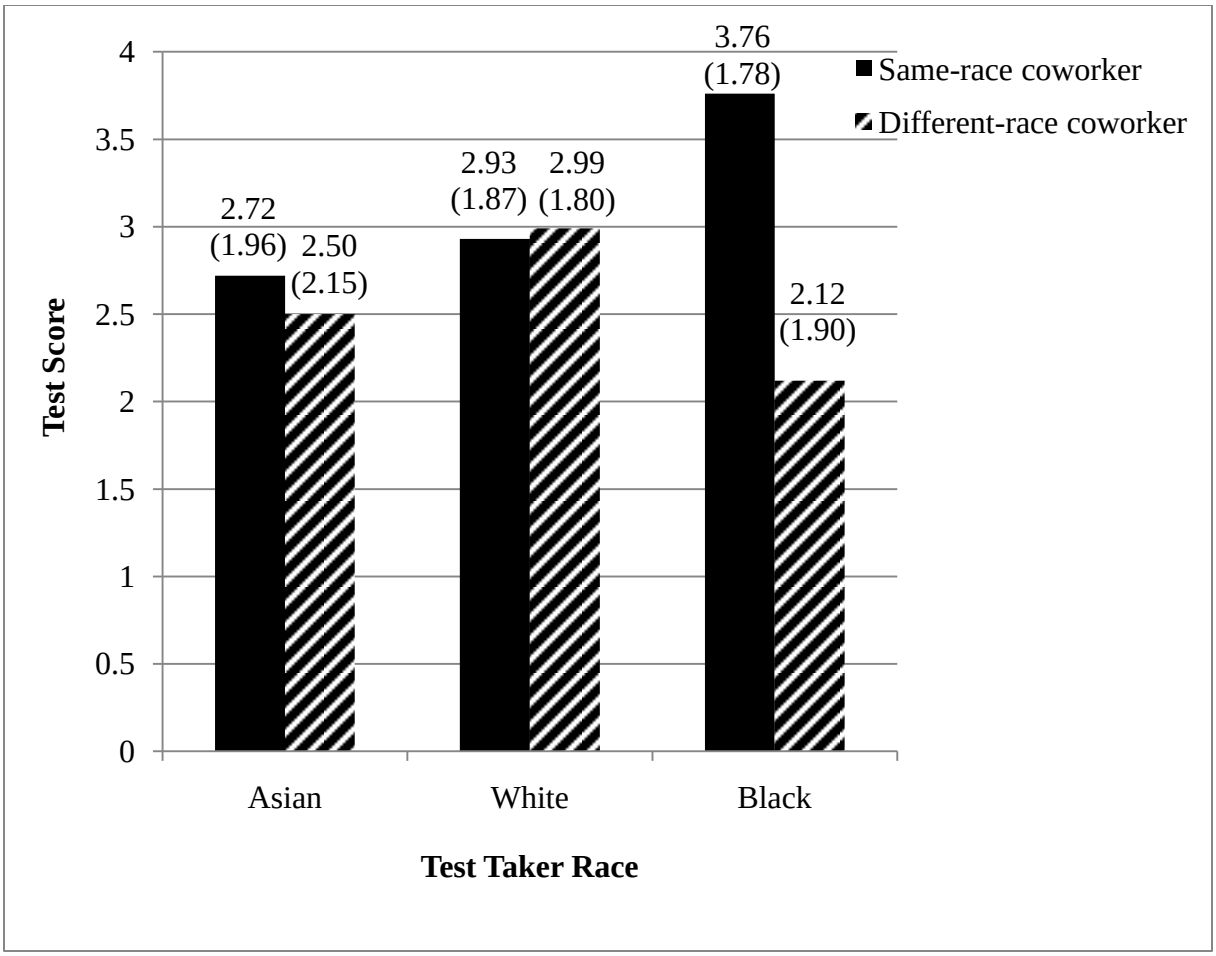

Figure 22. **Effects of race match with non-focal actors on test scores for different groups of respondents.**
Standard deviations are below the group means in parentheses.

Consistent with analyses reported earlier (in the context of testing hypotheses pertaining to helping and conflict) scores for Black respondents did not differ significantly on conflict item 1, $t(32) = -2.00$, $p = .054$, on conflict item 2, $t(32) = -1.37$, ns, and on helping item 1, $t(32) = -.44$, ns. As can be seen in Figure 23, the statistically significant difference was on helping item 2, $t(32) = -2.46$, $p < .05$. Black respondents got a higher score on average on this item (deciding how to respond to a coworker's request to switch shifts) when the coworker in the video was Black ($M = .76$, $SD = .66$) as opposed to White or Asian ($M = .18$, $SD = .73$).

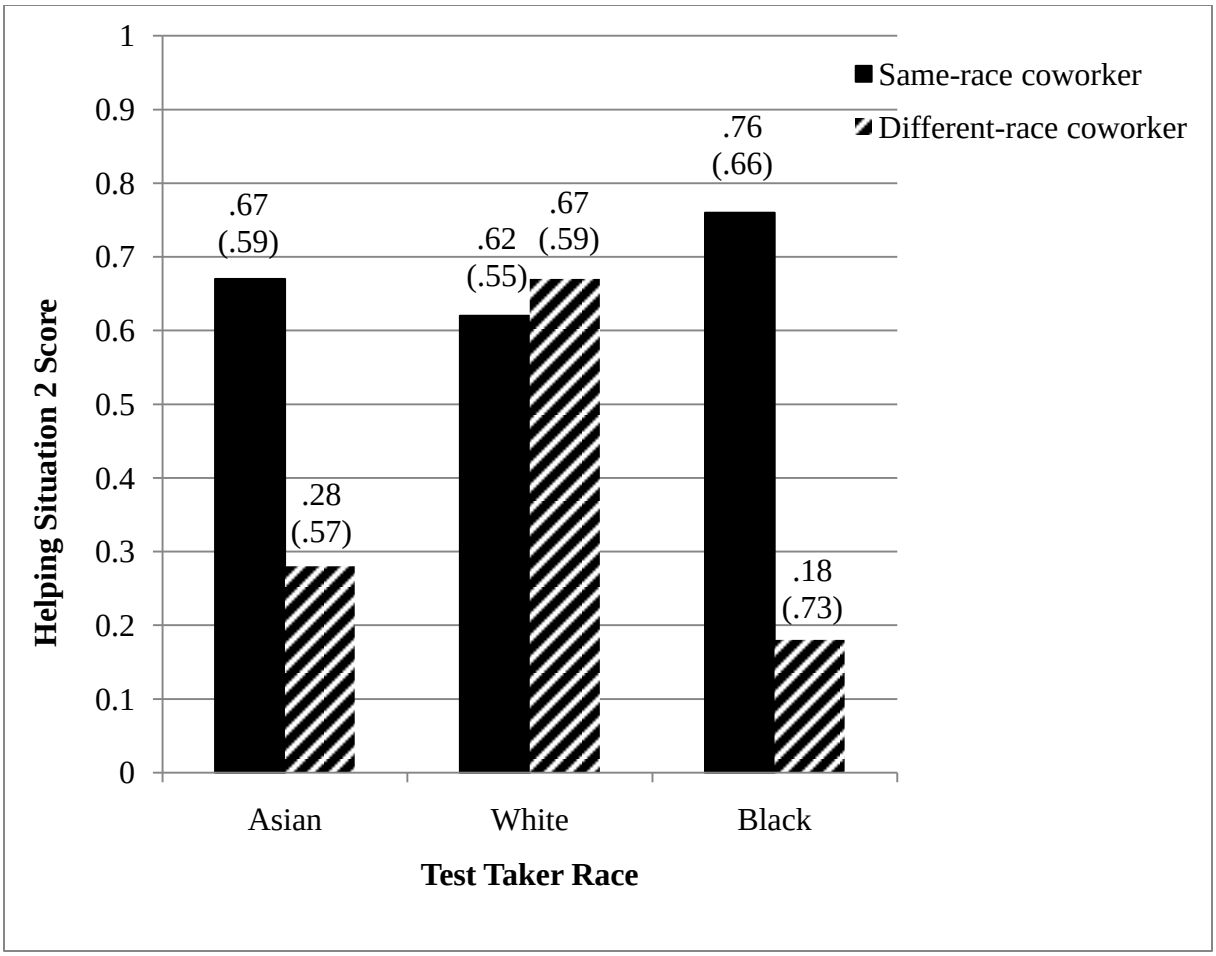

Figure 23. **Effects of race match with the non-focal actor on helping situation 2 scores for different groups of respondents.**
Standard deviations are below the group means in parentheses.

Recall that item scores were calculated based on most likely and least likely responses. Thus, to come up with a substantive explanation for the effect of race match on Black respondents' helping situation 2 scores, cross-tabs with a chi-square test of independence were calculated on their ML and LL responses to that SJT item. The goal was to examine how the frequencies with which different response options were selected as most or least likely were affected by Black respondents' race match with the non-focal actor in the video. Race match did not have a statistically significant effect on the frequency with which various response options

were selected as most likely, $\chi^2(4, N = 34) = 4.43$, ns (Fisher's exact test likewise did not support rejecting the null hypothesis of no effect for race match). Race match also did not have a statistically significant effect on the frequency with which various response options were selected by Black respondents as least likely, $\chi^2(4, N = 34) = 6.39$, ns (Fisher's exact test likewise did not support rejecting the null hypothesis of no effect for race match). Substantive analysis was therefore not justified.

Given the finding that Black respondents performed better on helping item 2 when the non-focal actor was Black, an analysis of mean differences between Blacks and Whites by race of the non-focal actor seems pertinent (see Table 26). Black test takers performed about as well as White test takers on the item in the Black non-focal actor condition, $t(94) = .33$, ns, but substantially worse ($M = .11$, $SD = .58$) than White test takers ($M = .67$, $SD = .56$) when the non-focal actor was not Black (i.e., White or Asian), $t(186) = -4.00$, $p < .001$. The effect size of this difference is large by Cohen's standards.

Table 26. **Helping item 2 score differences between Black and White test takers**

Condition	Test Taker Race	*M*	*SD*	*t*	*d*
Black non-focal actor	Black test taker (*n* = 13)	.77	.73	0.33	--
	White test taker (*n* = 83)	.71	.57		
Asian or White non-focal actor	Black test taker (*n* = 18)	.11	.58	-4.00**	-1.00
	White test taker (*n* = 170)	.67	.56		

Note. **$p < 0.001$.

Next, the findings in this study are discussed along with some of their practical implications. Study limitations and future directions for research are likewise considered. Before

proceeding, however, the results of the exploratory analyses in this section are summarized in Table 27 below.

Table 27. **Additional analyses of effects of demographic cues on performance**

	Finding				
Test	DV: SJT score	DV: Helping situation 1 score	DV: Helping situation 2 score	DV: Conflict situation 1 score	DV: Conflict situation 2 score
Is there a four way interaction of respondent race, respondent gender, race match between respondent and non-focal actor, and gender match between respondent and non-focal actor?	No; two-way interaction of respondent race and race match between respondent and non-focal actor	--	--	--	--
Which item(s) explain the two-way interaction of respondent race and race match on total SJT score (i.e., which item(s) do Black test takers perform differently depending on race match)?	--	No	Yes	No	No
Do Black test takers perform worse than White test takers when the non-focal actor is Black?	--	--	No	--	--
Do Black test takers perform worse than White test takers when the non-focal actor is not Black?	--	--	Yes	--	--

DISCUSSION

The purpose of this study was to consider the implications that demographic cues in video-based situational judgment tests may have for how test takers of different demographic profiles (gender, racial) respond to the SJT items as well as how they react to the test. Two types of items, conflict and helping, were investigated. Hypotheses based on research on relational demography, conflict, helping, and applicant reactions were tested.

Study Findings

Helping. Almost no support was found for hypotheses pertaining to how test takers would respond to helping situations depending on their own as well as the non-focal actors' race and gender. The expectation that the effects of race mismatch on helping would be more negative for White respondents viewing a minority non-focal actor than for minority respondents viewing a White non-focal actor, based on relational demography research showing asymmetrical effects of demographic similarity for high and low status groups, was not met. There was also no support for hypotheses, likewise based on asymmetrical effects found in relational demography research, about particular respondent groups offering specific non-focal actors lower levels of help relative to levels of help under other combinations of respondents with targets. Also, Asians, the "model minority" (Alvarez et al., 2006), who are stereotypically believed to have high work ethic, were not found to receive the highest level of help, in contrast to expectations. There was partial support for the hypothesis that female respondents will offer more help to non-focal actors than will male respondents, and this finding fits in with research showing that women, as compared to men, tend to perform more "altruistic" behaviors targeted toward individuals needing assistance (Eagly, 2009).

Conflict. Hypotheses pertaining to conflict received almost no support. It was not the case that respondents endorsed more assertive and cooperative responses when seeing a non-

focal actor of the same race. This expectation had been based on findings that groups will endorse a conflict avoidance strategy less so when considering conflict with those of the same race (Davidson, 2001). Likewise contrary to expectations, respondents' race and race match with the non-focal actor did not interact to predict respondents' preference for third party intervention in conflict situations. Surprisingly, there were no differences in cooperativeness and assertiveness shown by male and female respondents, despite studies on gender norm expectations that would suggest differences should be seen (e.g., Graves & Elsass, 2005; Wilson et al., 2003). On one response option, both men and women tended to be less assertive and more cooperative in response to female non-focal actors. This support for my hypothesis, though minimal, fits in with findings of benevolent sexism (Wilson & Gallois, 1985) and more positive behavior (Carli, 1989) toward women. There was no support for hypotheses about certain groups responding to specific non-focal actors with higher or lower levels of assertiveness, cooperativeness, or preference for intervention relative to observed levels given other combinations of respondents with targets. Finally, respondents' gender and gender match with the non-focal actor interacted to predict preference for third party intervention in conflict situations, but only in the case of one response option (both male and female respondents were less willing to get a third party to intervene in an unglued cartons situation when the non-focal actor was female).

Examining reasons for lack of support for hypotheses. Although only a small set of the helping and conflict hypotheses were supported (and in most cases just partially so), this is due in part to the inability to test most of these at the level of dimensions of responses as had been planned. One of the hypotheses that could be tested at the dimension level, that male respondents would be more willing to help female non-focal actors than male ones while female

respondents would not distinguish between male and female actors when offering help, stands out as having been supported. This finding is very much in line with meta-analytic findings on the role of gender in helping (Eagly & Crowley, 1986).

It should be acknowledged, however, that there remains the possibility that most of the hypotheses would still not have been supported even if they could have been tested as planned. Ideas from relational demography may have limited application to the context of test taker responses on a video-based SJT. Relational demography takes into consideration the different profiles of those who interact with one another to make predictions of outcomes for the interacting parties. The theory was not completely applied in the case of this research in the sense that the focal actor in each video was left out of the equation. It is possible that respondents took not only their own and the non-focal actor's demographics into consideration when responding to the test question, but also incorporated information on the demographics of the focal actor (White female) into their decisions of how to respond. It cannot be ruled out, for example, that some respondents actually assumed the role of the focal actor in the situation, taking on her demographic profile, to respond to a given situation, or that some respondents, without actually assuming the focal individual's role, simply provided what they believed to be a desirable response from the standpoint of a White female. This research may have produced more significant findings if it had been possible to create situations that featured just one actor, while only implying the presence of the other one (see MacDorman, Coram, Ho, & Patel, 2010 for an example). In that case, there would be no physically present "focal" actor to take into consideration. The relational demography perspective could then be more precisely mapped onto the "interaction" of the respondent and actor.

Applicant reactions. Surprisingly, the findings pertaining to applicant reactions were mostly inconsistent with research suggesting that minorities (e.g., Blacks, women) react positively to diversity (e.g., Avery et al., 2004). The only exception to this was that minorities (Asians, Blacks) did perceive better opportunity to perform when they viewed minority non-focal actors. This limited support of my reactions hypotheses is further discussed in the limitations section.

Response effectiveness. Arguably, the most interesting findings in this study were the more exploratory ones pertaining to response effectiveness (i.e., actual performance on the test). As a group, minority and White respondents performed somewhat worse on helping item 1 when the non-focal actor did not match them in race. Only minority respondents performed worse on helping item 2 when the non-focal actor did not match them in race. It was because of this particular item that Black respondents scored higher on the SJT when the non-focal actor was Black as opposed to White or Asian. There was also a significant three-way interaction of respondent sex, non-focal actor sex, and non-focal actor race on helping item 1 score. With a White non-focal actor (but not a Black one), men and women scored differently depending on the actor's gender. Male and female respondents got equal scores on helping item 1 when the White non-focal actor was female. But when the White non-focal actor was male, female respondents scored higher on the item than male respondents. This three way interaction was not observed on helping item 2 though. Finally, minority respondents performed better on both conflict items when the non-focal actor was female, while White respondents were not influenced by non-focal actor gender in these items.

The theoretical background for this study suggests that differences in item responses may occur based on demographic cues in the test. However, a substantive analysis of the link between

item scores and most and least likely responses to these items (i.e., the basis for determining item scores) could not be done in most cases. This makes it difficult to interpret substantively the effects that were found. Furthermore, some of the inconsistency in effects found across the same type of items (i.e., helping) makes it even harder to offer a theoretical explanation. It should be noted, as well, that because the items were not written for the constructs of interest, any theory would have to be imposed on the items. Larger sample sizes would have facilitated substantive analyses of score differences. However, the limited ability to offer theoretical interpretations for some of the findings should not eliminate their practical value.

Practical Implications

The finding that Black test takers scored lower on average than White test takers on helping item 2 when the non-focal actor was Asian or White, but not when the non-focal actor was Black, has special practical relevance. It suggests that test developers potentially have some amount of control over test bias when designing video-based tests. Yet, this issue needs to be further investigated to determine the extent to which the current finding is generalizable to other types of items.

Given the findings in this study it is appropriate to ask how large an issue for consideration demographic cues in video based tests poses for test developers. The appropriate answer is arguably "it depends." Current findings do show demographic cues at times having an influence both on responses and actual performance on the test, as well as on minority respondents' perceived opportunity to perform. The extent to which companies may be compelled to take these initial findings into consideration in developing video-based test materials may depend on factors such as the diversity of their applicant pool (e.g., does the company have a lot of difficulty hiring enough minorities?) and importance placed on projecting

a positive image of inclusion to minority applicants (i.e., does the company make it a goal to convey organizational diversity?). The competencies assessed by a video-based test would also be important to consider. The current findings speak to demographic cue effects that may occur for interpersonal types of situations having to do with conflict and helping between coworkers. Any results are therefore of more relevance to companies assessing these types of competencies.

Limitations and Future Directions

Although this study makes significant inroads into an area of study that past research has offered little insight into, and provides some interesting initial findings with regard to the impact of demographic cues in video-based situational judgment tests on responses to, performance on, and reactions to these types of tests, it does have a number of limitations that should be discussed. Some of these have to do with sampling, inherent multidimensionality of situational judgment tests, context and administration factors, the set of variables that were chosen for inclusion in (versus exclusion from) this research, and limited sampling from the domain of SJT item types.

Sample size. Although a lot of time and effort was invested into getting enough men and minorities for this study, this effort was more successful with regard to the former than the latter demographic. Numbers of Asian and Black respondents were relatively small in this study and particularly in the case of the four-way interactions I tested, there were really too few to find significant effects even if these effects were there to be found. Thus, I have little confidence in the null findings with regard to any of the four-way interactions. This area of research would yield well to studies that have much larger numbers of minorities to sample from.

Multidimensionality. I tried to discover theoretical dimensions in responses to the SJT items. However, these items were not designed to measure the dimensions I tried to pull out. As

described earlier, item stems were based on critical incidents technicians provided and response options were actions these technicians proposed as effective or ineffective ways of responding to the incidents. Given this non-theoretical approach to test development, it is not surprising that the dimensions I hoped to find were not reliably measured by the SJT response options. The multidimensional nature of the SJT items in this study meant that I could not test my hypotheses about *dimensions of responses* as planned. Even though several of the hypotheses about helping as a dimension could be tested, this helping score was ultimately composed of just a fourth of the response options I meant for it to represent. Thus, most hypotheses had to be tested at the level of the response options instead of dimensions and my ability to provide a compelling explanation for why respondents may have endorsed a particular option more or less strongly depending on their own race/gender, or an interaction of their own demographics with demographic cues in the video, was limited. It should be pointed out, however, that the problem of failing to derive construct validity for the SJT items in my study is not unique, but typical to SJT research. Most SJTs are developed via non-theoretical approaches and little attention is paid to constructs assessed by these tests (Christian, Edwards, & Bradley, 2010 provide a review). Though it is likely to be difficult, future research can investigate the effects of demographic cues on SJT items developed via a theoretical approach so as to have the ability to draw stronger theoretical conclusions from any findings.

Weights. Response item weighting in this study may be argued to be a limitation. Even though I was able to establish adequate agreement among judges to use the means of ratings they provided as weights for likelihood ratings respondents assigned to response options, the adequate agreement came with the collapsing of the five-point scale judges used, into a three-point scale. Even utility of the practice of weighting by means can be questioned as many will argue that unit

weights would have been adequate (see Bobko, Roth, & Buster, 2007 for a discussion). As I also tried to apply unit weights as an alternative to weighting by means and the results of analyses ended up about the same, the approach did not make a difference in this study. Ultimately, weighting was done for the sake of one dimension of one SJT item so the issue of weights actually has limited bearing on the findings in this study.

Context and administration. This study examines in a lab setting a type of test that in practice would be administered in a selection context. While the experimental nature of this study is a strength in that variables of interest could be isolated in the lab, it is simultaneously a weakness in that high stakes testing is associated with generally higher test taker motivation and effort than typically seen in a lab setting. Further, SJTs, including the one items for this study were drawn from, are typically administered as timed tests. For the original test respondents had 70 seconds to answer an item (and there was a ten second notification built in). As the 4-item test in my study was not timed, respondents had as much time as they wanted to deliberate before providing their responses. It is not possible to know how the results of this study may have been different given a timed selection context where test takers would be more motivated and forced to make faster decisions. A direction for future research may be to consider the moderating effects that motivational and administration-related factors (e.g., time, response instructions) could have on the relationship of demographic cues in SJTs with test takers' responses.

The lab context may have also contributed to the limited findings with regard to applicant reactions. Respondents might have perceived the instructions to "think of yourself as a job seeker applying for a technician job with ABC Corporation," as too artificial. They may also not have recognized the connection between the job description for a plant technician and the interpersonal content of the test they were taking "to see if you will be further considered by

ABC Corporation for a technician position," although an attempt was made to make that connection clear in the instructions (which respondents may or may not have fully read). If the respondents perceived the reactions questions to lack relevance, this overarching impression may have driven their responses to these questions more so than demographic cues in the test. On average, test takers responded somewhere between "somewhat disagree" and "neither disagree nor agree" to scales for opportunity to perform, and job relatedness (both predictive and content). They averaged at about "neither disagree nor agree" on the compatibility scale and between "neither disagree nor agree" and "somewhat agree" on the attraction scale. These results do seem to suggest that respondents may have perceived the questions to be of limited relevance, and therefore did not have particularly strongly reactions to them. Future research may add to current findings by making an effort to maximize ecological validity when investigating applicant reactions to demographic cues in tests.

Categorization process. In this study, hypotheses predicting effects of demographic cues on test takers' response tendencies assumed that a categorization process occurs for respondents when they view actors of a certain race and gender. Based on the finding that respondents sometimes were not able to correctly identify the demographic profile(s) of the actor(s) they viewed, there is reason to question whether all respondents were able to categorize actors into their "correct" demographic groups. Aside from reasons of inattention or purposeful error, there are two broad possibilities as to why these categorization errors may have occurred. The first possibility is that some respondents, though able to correctly categorize the actors mentally, were later unable to assign correct labels to these actors when asked (e.g., some may have been confused about what the label Middle Eastern actually means) or were unwilling to do so (e.g., recent media reports have discussed contemporary youths' tendency toward avoiding placing

themselves or others into narrow demographic categories; Saulny, 2011a; 2011b). The second possibility is that some respondents were unable to even categorize the actors mentally. This latter scenario presents a bigger problem as it has the categorization process actually not occurring. It is not possible to know the extent to which the test takers in this study experienced one or the other problem. Yet, as respondents who failed the manipulation check(s) were excluded from all analyses, as were multiracial and "other" respondents from analyses pertaining to race, the influence of this issue of categorization failure on the results of this study should have been minimized. An interesting direction for future research may be to consider the implications that changing trends in demographic categorization of self and others has for how test takers perceive and respond to others in the context of selection testing (both in terms of test responses and applicant reactions).

Also related to the issue of categorization, it should be noted that Asian respondents in this study were a much less homogenous group than the White and Black respondent groups. For example, both East Asians and Indians were part of the "Asian" group of respondents in this study. While Indian respondents most likely did not see the East Asian non-focal actors in the videos as belonging to the same demographic group as them, this distinction could not be made in the analyses pertaining to non-focal actor race and race match between respondent and non-focal actor, as respondents were not asked to report what region of the Asian continent they associate themselves with. Those analyses may be muddied then and perhaps true effects of racial cues for *East* Asian respondents were masked because of this. It would be important for future research in this area to address this limitation by better identifying when respondents are actually paired with non-focal actors of the same demographic profile as them. Such research

would be able to shed better light on whether Asian respondents are sensitive to demographic cues in video-based tests.

Limited demographic profiles of actors. To keep the design of this study manageable, decisions had to be made regarding which demographic cues to focus on. The decision was made to keep the age of the actors constant and to limit non-focal actor race to three groups only. As a result, this study cannot offer findings regarding how age may interact with other demographic cues to influence responses to the SJT and how respondents may react to Hispanic actors in situational judgment items, for example. This limits the practical implications of this study as organizations are likely to develop SJTs that show a wider range of ages and races (so as to convey diversity) than I examined. Other research can examine additional demographic cues as predictors of test takers' responses to SJTs. As this study also kept the race, gender, and age of the *focal* actors constant, investigating possible effects of focal actor (i.e., again, the one the test taker needs to assume the role of) demographics on test taker outcomes is likewise a topic for future research to consider.

Limited demographic profiles of respondents. Respondents in this study were less diverse in terms of their demographics (e.g., age, race, education, job experience) than is the typical organization's applicant pool. The findings of this study do not address, for example, whether more experienced test takers may respond similarly to demographic cues in SJTs with regard to response tendencies and actual performance. It is possible, for instance, that those with relevant job experience will be more likely to respond effectively to SJT questions regardless of the demographic cues that may be present in the test. Future research can look into these types of questions.

Item type and test length. The generalizability of the current findings (and in certain cases lack thereof) may be limited by the fact that, due to resource limitations, only 4 items were included in this study. Further, they were just two types of interpersonal items (i.e., conflict, helping), albeit very common ones, out of a wider range of possible SJT item content types (e.g., handling a technical work problem, delivering bad news to a subordinate, providing encouragement to a coworker). The findings, particularly the practical ones with regard to performance differences of Black test takers depending on race cues in the test, may or may not generalize to other types of items and longer tests. This question in particular would be important to investigate further given the important dilemma of adverse impact in high stakes testing and its implications for diversity in hiring.

Overview. Despite the limitations in this research study, it makes an important contribution to the literature by applying a relational demography perspective to examining responses to a video-based SJT. This study adds to researchers' knowledge of the relevance relational demography has for issues in selection (e.g., recruitment; Avery, Hernandez, & Hebl, 2004; Goldberg, 2005; Perkins, Thomas, &Taylor, 2000; interviews; Buckley et al., 2007). It fits in well with the extensive research that examines sources of construct-irrelevant variance in SJT performance (e.g., the extent to which characters look/act human; MacDorman et al., 2010; administration medium; Lievens & Sackett, 2006; faking; Peeters & Lievens, 2005; response instructions; Nguyen & McDaniel, 2003; Ployhart, & Ehrhart, 2003) as well as other types of tests (e.g., cognitive ability; Brown & Day, 2006; Schmitt, 1988; personality; Ellingson, Sackett, & Hough, 1999; math/mechanical ability; Grand, Ryan, Schmitt, & Hmurovic, 2011; Hofstetter, 2003). This cited line of research typically considers the extent to which, given the presence of construct irrelevant variance, tests still predict outcomes of interest. Examining this issue was

beyond the scope of the current research. This would, however, be a useful direction for the current research to go in. Thus, given that demographic cues were found to impact test performance in some cases, it would be important to examine how the criterion-related validity of video-based SJTs may vary depending on the interaction of test characteristics (actor race, gender) with test taker characteristics (respondent race, gender). As discussed earlier, as research moves forward, a broader set of test and test taker characteristics might be useful to consider than was included in the current study.

Future research should continue to apply a relational demography perspective to video-based, and other types of SJTs that use a visual medium, where test takers have the opportunity to view the actors or characters involved. The newest trend in SJT development is the use of avatars instead of human actors. Some of the same relationships examined in the current study would be relevant to examine using avatars as they, like human actors, have demographic profiles. The use of avatars may be a particularly promising direction for SJTs with regard to test fairness in that avatar SJTs are more amenable than human actor SJTs to customization according to the needs or preferences of the test taker (i.e., there is no need to recreate the test if changing a character's race/gender/age is desired). Research will need to examine this new medium further to determine how test taker performance and reactions may be optimized via customization. This direction of research would have a lot of practical value for test developers and organizations.

Conclusion

This study applied a relational demography perspective to examine differential effects of demographic cues in a video-based situational judgment test on test takers belonging to different race and gender groups. Findings suggest some utility for looking at test performance and

reactions from the standpoint of the interaction of respondents' and actors' demographic profiles. Further research looking into ways in which actor-related elements of video-based test design may be used to enhance particular test taker groups' performance and reactions seems warranted.

APPENDICES

APPENDIX A

Text Preceding SJT Test (Adapted from Bauer et al., 2001)

Please think of yourself as a job seeker applying for a technician job with ABC Corporation. This company is offering a yearly salary 20% higher than other companies you're considering and is located in town you like. In talking with people hired in the last 5 years, you have discovered that employees have received an average of three promotions in that time. The company also has been rated as a leader in the industry in terms of proactive environmental policies and was rated as one of the top 100 places to work by *US News & World Reports*.

Today you will be taking an initial screening test to see if you will be further considered by ABC Corporation for a technician position within their company. It is a test of interpersonal ability. You will be viewing and responding to a number of scenarios reflecting interpersonal situations that may be encountered on the technician job. For each scenario you view, you will be asked to provide your most as well as least likely course of action in that situation.

After the test you will be filling out several short questionnaires. Thank you for your participation.

APPENDIX B

Job Description for an Entry Level Plant Technician

Plant technicians are hired at entry level as members of a work team. Technician work teams are responsible for keeping production flowing, for meeting performance goals and for maintaining the high level of quality that consumers expect from the company's products.

Technicians do various types of work, including running high-speed converting and packing equipment, controlling the process from a computer terminal, and performing preventive maintenance on equipment. Technicians are expected to get involved, take charge of situations and confront business-related problems.

Technicians generally have strong technical, interpersonal and problem-solving skills.

The typical plant operates 7-days per week. Technicians work any day of the week, including weekends. The hours of work vary from 8 to 12 hours per day. Shift rotations vary from working a single straight shift to working two or three rotating shifts. Technicians must be willing to rotate and to work any shift.

APPENDIX C

Situational Judgment Test Items

Note:

1. The "Other Person" was one of three races (White, Black, or Asian) and either male or female depending on the study condition. Response options referred to the Other Person as "he" and "him" when that actor was male and as "she" and "her" when that actor was female. The term "Other Person" was not used during the experiment.

2. "Focal Person" was always a White female. The Focal Person in each scenario had a name (the names were Alice, Jane, Carol, and Ann); the term "Focal Person" was not used during the experiment.

Conflict Situation 1 (Notes: Competence stereotype is relevant; goal of intervention would be to report coworker's poor performance)

Focal Person: What's going on? This is the fourth one in a row that hasn't been glued shut. Oh, I better call over.

Focal Person (on phone)**:** Yeah, this is Alice in shipping. Yeah, you guys are sending over cartons that aren't glued shut. So, you better shut down that line and figure out what your problem is.

Other Person: Hey, I'm standing here looking at the line and every case is perfect. Our case packer is working just fine.

Please rate each response option according to how likely you would be to choose that response if you were Focal Person.

If you were Focal Person, what would you MOST and LEAST likely do?

a. Tell your coworker that you will continue to watch for unglued cartons and get back to him (her).

b. Assume that the cartons are being glued properly because your coworker hasn't noticed the problem.

c. Insist that there must be a problem your coworker hasn't noticed since unglued cartons are arriving at the warehouse.

d. Ask another team member to discuss the problem with your coworker.

e. Let your team leader know that your coworker is allowing unglued cartons to get to the warehouse.

Help Seeking Situation 1 (Notes: Competence stereotype is relevant)

Other Person: Carol, can I bug you for a minute?

Focal Person: Yeah, what?

Other Person: I'm having a real problem with Frank. I don't think he listens to me. He doesn't think I know anything. I was hoping you might give me some ideas on how to deal with him better?

Please rate each response option according to how likely you would be to choose that response if you were Focal Person.

If you were Focal Person, what would you MOST and LEAST likely do?

a. Explain to your coworker that you are very busy, but if he (she) wants to talk later, you'll tell him (her) how you handle Frank.

b. Suggest that your coworker ignore Frank because working with Frank can be difficult.

c. Tell your coworker that he (she) needs to learn how to handle his (her) own conflicts with Frank.

d. Tell your coworker how you deal with Frank.

e. Suggest that your coworker meet with Frank to talk about their differences.

f. Suggest that your coworker talk to the team leader about his (her) problem with Frank.

Conflict Situation 2 (Notes: Stereotypes pertaining to aggressiveness and agreeableness are relevant; goal of intervention would be reconciliation)

Focal Person: Hey, can I sit here for a minute.

Other Person: Sure. Sit there all day if you like. I was just leaving.

Focal Person: Wait a minute. I want to find out why you're always giving me a hard time.

Other Person: Look, I only give people a hard time if they make mistakes. Just don't make mistakes and we'll get along fine.

Please rate each response option according to how likely you would be to choose that response if you were Focal Person.

If you were Focal Person, what would you MOST and LEAST likely do?

a. Ask your coworker what you could do to get along better.

b. Ask the team leader for suggestions about how to work with your coworker.

c. Try to avoid your coworker whenever possible because you can't change him (her).

d. Tell your coworker that he (she) should act like a team member and try to work together.

e. Ask the team leader to meet with you and your coworker to work out your differences.

f. Ask other team members for suggestions on how to work with your coworker.

<u>Help Seeking Situation 2</u> (Notes: Work ethic stereotype is relevant)

Other Person: Hey Ann, can you do me a real big favor? I want to take a training class next week but I'm on night shift. Can you switch with me?

Focal Person: I took nights for you like a month ago so you could take some training class, didn't I?

Other Person: Yeah, but this is one the new I-3000 wrapping equipment and I think it would be real helpful. Come on, can you switch with me again?

Focal Person: Hey, I don't even think you've used the training you got the last time. And quite frankly, I don't like working nights anymore than you do. And it seems to me like your always coming with some new idea on how not to work nights, man.

Other Person: Listen, we're supposed to help out members of our team, right? I already got permission to sign-up and you've already had this training. I think you should give me the chance to learn something that's going to help us do a better job.

Please rate each response option according to how likely you would be to choose that response if you were Focal Person.

If you were Focal Person, what would you MOST and LEAST likely do?

a. Agree to switch with your coworker so that he (she) can take the training course.

b. Agree to switch with your coworker, but only if he (she) will switch with you in a few weeks.

c. Tell your coworker you think he (she) is taking advantage of the system and you won't switch.

d. Tell your coworker you won't switch, and he (she) should wait until training is offered at another time that doesn't interfere with his (her) schedule.

e. Tell your coworker that you think he (she) is taking advantage of the system, but you will help him (her) find someone else to switch.

APPENDIX D

Reactions Items Pertaining to Opportunity to Perform (Adapted from Bauer et al., 2001)

1. I could really show my skills and abilities through this test.
2. This test allowed me to show what my interpersonal skills are.
3. This test gives individuals the opportunity to show what they can really do.
4. I was able to show what I can do on this test.

APPENDIX E

Reactions Items Pertaining to Job-Relatedness (Adapted from Bauer et al., 2001)

Job Relatedness—Predictive
1. Doing well on this test means a person can do the technician job well.
2. A person who scored well on this test will be a good technician.

Job Relatedness—Content
1. It would be clear to anyone that this test is related to the technician job.
2. The content of the test was clearly related to the technician job.

APPENDIX F

Reactions Items Pertaining to Attraction (Adapted from Perkins et al., 2000)

1. I would request additional information regarding the possibility of employment with this company.
2. I would like this company to recruit on campus.
3. I would speak to a company representative about the possibility of employment.
4. I would think this organization is attractive.
5. I would not recommend this company to a friend. (Reverse coded)
6. I would like this organization.

APPENDIX G

Reactions Items Pertaining to Compatibility (Adapted from Perkins et al., 2000)

1. I would feel at home working for an organization like this.
2. I would very much like to work for this organization.
3. This organization will likely meet my desires and needs.
4. I would have no problems adjusting to this organization.

APPENDIX H

Demographic Questions

1. What is your age? _____

2. What is your gender?
a. Male
b. Female

3. Is your ethnicity Hispanic/Latino?
a. Yes
b. No

4. If your ethnicity is not Hispanic/Latino please select your race below. Select more than 1 response if applicable.
a. American Indian or Alaska Native
b. Asian
c. Black or African American
d. Native Hawaiian or Other Pacific Islander
e. White
f. Other

5. If other, please specify. _____

6. What is your class standing?
a. Freshman
b. Sophomore
c. Junior
d. Senior
e. Other/non-degree

7. What is your cumulative college GPA?
a. less than 2.00
b. 2.00 to 2.29
c. 2.30 to 2.59
d. 2.60 to 2.89
e. 2.90 to 3.19
f. 3.20 to 3.49
g. 3.50 to 3.79
h. 3.80 to 3.99
i. 4.00

8. Which of the following best describes your intended or current major?
a. None, I have not yet decided on a major
b. Business
c. Engineering

d. Fine Arts or Humanities
e. Social Science (i.e., Psychology, Sociology, Political Science, etc)
f. Natural Science (i.e., Biology, Chemistry, Physics, etc)
g. Other

APPENDIX I

Distributions of Responses for Those Who Failed Manipulation Checks

Table 28. **Distribution of responses for those who failed the manipulation check for conflict situation 1**

Condition	East Asian Female	East Asian Male	Black Female	Black Male	Caucasian Female	Caucasian Male	Middle Eastern Female	Middle Eastern Male	Cannot tell	Total *N*
Asian Female non-focal	X	--	--	--	--	--	100%	--	--	4
Asian Male non-focal	--	X	--	--	--	--	--	50%	50%	2
Black Female non-focal	--	--	X	50%	--	--	--	--	50%	2
Black Male non-focal	--	7.7%	--	X	--	--	--	76.9%	15.4%	13
White Female non-focal	--	--	--	--	X	--	--	--	100%	4
White Male non-focal	--	--	--	--	75%	X	--	--	25%	4

Note. X indicates the correct response to the manipulation check for conflict situation 1 given a particular condition.

Table 29. **Distribution of responses for those who failed the manipulation check for conflict situation 2**

Condition	East Asian Female	East Asian Male	Black Female	Black Male	Caucasian Female	Caucasian Male	Middle Eastern Female	Middle Eastern Male	Cannot tell	Total *N*
Asian Female non-focal	X	--	--	--	--	--	50%	--	50%	6
Asian Male non-focal	--	X	--	--	--	--	--	66.7%	33.3%	3
Black Female non-focal	--	--	X	--	--	--	--	--	--	0
Black Male non-focal	--	--	--	X	--	--	--	--	--	0
White Female non-focal	12.5%	--	--	--	X	--	--	--	87.5%	8
White Male non-focal	--	9.1%	--	--	--	X	--	--	90.9%	11

Note. X indicates the correct response to the manipulation check for conflict situation 1 given a particular condition.

Table 30. **Distribution of responses for those who failed the manipulation check for helping situation 1**

Condition	East Asian Female	East Asian Male	Black Female	Black Male	Caucasian Female	Caucasian Male	Middle Eastern Female	Middle Eastern Male	Cannot tell	Total *N*
Asian Female non-focal	X	100%	--	--	--	--	--	--	--	1
Asian Male non-focal	50%	X	--	--	--	--	--	25%	25%	4
Black Female non-focal	--	--	X	--	--	--	--	--	--	0
Black Male non-focal	--	--	--	X	--	--	--	--	100%	1
White Female non-focal	20%	--	--	--	X	--	--	--	80%	5
White Male non-focal	--	--	--	--	--	X	--	--	100%	1

Note. X indicates the correct response to the manipulation check for helping situation 1 given a particular condition.

Table 31. **Distribution of responses for those who failed the manipulation check for helping situation 2**

Condition	East Asian Female	East Asian Male	Black Female	Black Male	Caucasian Female	Caucasian Male	Middle Eastern Female	Middle Eastern Male	Cannot tell	Total *N*
Asian Female non-focal	X	--	--	--	--	--	66.7%	--	33.3%	3
Asian Male non-focal	--	X	--	--	--	--	--	66.7%	33.3%	3
Black Female non-focal	--	--	X	--	--	--	100%	--	--	1
Black Male non-focal	--	--	11.1%	X	--	--	--	33.3%	55.6%	9
White Female non-focal	--	--	--	--	X	14.3%	--	--	85.7%	7
White Male non-focal	--	--	--	--	75%	X	--	--	25%	4

Note. X indicates the correct response to the manipulation check for helping situation 2 given a particular condition.

REFERENCES

REFERENCES

Alvarez, A. N., Juang, L., & Liang, C. T. H. (2006). Asian Americans and racism: When bad things happen to "model minorities." *Cultural Diversity and Ethnic Minority Psychology, 12*, 477-492.

Auster, C. J., & Ohm, S.C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles, 43*, 499-528.

Avery, D. R. (2003). Reactions to diversity in recruitment advertising—Are differences Black and White? *Journal of Applied Psychology, 88*, 672-679.

Avery, D. R., Hernandez, M., & Hebl, M. R. (2004). Who's watching the race? Racial salience in recruitment advertising. *Journal of Applied Social Psychology, 34*, 146-161.

Avery, D. R., & McKay, P. F. (2006). Target practice: An organizational impression management approach to attracting minority and female job applicants. *Personnel Psychology, 59*, 157-187.

Banki, S., & Latham, G. P. (2010). The criterion-related validities and perceived fairness of the situational interview and the situational judgment test in an Iranian organisation. *Applied Psychology: An International Review, 59*, 124-142.

Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment, 12*, 135-148.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology,54*, 387-419.

Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14*, 223-235.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. *Advances in Experimental Social Psychology, 7*, 157-215.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods,10*, 689-709.

Brewer, M. B., & Harasty Feinstein, A. S. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken, & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 255–270). New York: Guilford Press.

Brooks, M. E., & Highhouse, S. (2006). Can good judgment be measured? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 39-55). San Francisco: Jossey-Bass.

Brown, R. P., & Day, E. A. (2006). The difference isn't Black and White: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*, 979-985.

Browne, I., & Misra, J. (2003). The intersection of gender and race in the labor market. *Annual Review of Sociology, 29*, 487-513.

Brumbaugh, A. M. (2009). Why do I identify with thee? Let me count three ways: How ad context influences race-based character identification. *Psychology & Marketing, 26*, 970-986.

Buckley, M. R., Jackson, K. A., Bolino, M. C., Veres, J. G. III., & Feild, H. S. (2007). The influence of relational demography on panel interview ratings: A field experiment. *Personnel Psychology, 60*, 627-646.

Carli, L. L. (1989). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology, 56*, 565-576.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233-254.

Chatman, J. A., & Flynn, F. J. (2001). The influence of demographic heterogeneity on the emergence and consequences of cooperative norms in work teams. *The Academy of Management Journal, 44*, 956-974.

Chatman, J. A., & O'Reilly, C. A. (2004). Asymmetric reactions to work group sex diversity among men and women. *Academy of Management Journal, 47*, 193-208.

Chattopadhyay, P., Tluchowska, M., & George, E. (2004). Identifying the ingroup: A closer look at the influence of demographic dissimilarity on employee social identity. *Academy of Management Review, 29*, 180-202.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83-117.

Dalessio, A. T. (1994). Predicting insurance agent turnover using a video-based situational judgment test. *Journal of Business and Psychology, 9*, 23-32.

Davidson, M. N. (2001). Know thine adversary: The impact of race on styles of dealing with conflict. *Sex Roles, 45*, 259-276.

De Dreu, C. K. W., Evers, A., Beersma, B., Kluwer, E. S., & Nauta, A. (2001). A theory-based measure of conflict management strategies in the workplace. *Journal of Organizational Behavior, 22,* 645-668.

Deshpandé, R., & Stayman, D. M. (1994). A tale of two cities: Distinctiveness theory and advertising effectiveness. *Journal of Marketing Research, 31*, 57-64.

Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist, 64*, 644-658.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*, 283-308.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin, 117*, 125-145.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.

Elsass, P. M., & Graves, L. M. (1997). Demographic diversity in decision-making groups: The experiences of women and people of color. *Academy of Management Review, 22,* 946-973.

Ensher, E. A., & Murphy, S. E. (1997). Effects of race, gender, perceived similarity, and contact on mentor relationships. *Journal of Vocational Behavior, 50*, 460-481.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model, ten years later. In S.Chaiken, & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231–254).New York: Guilford Press.

Foley, S., Linnehan, F., Greenhaus, J. H., & Weer, C. H. (2006). The impact of gender similarity, racial similarity, and work culture on family-supportive supervision. *Group & Organization Management, 31*, 420-441.

Gaines, S.O., Jr., Marelich, W.D., Bledsoe, K. L., Steers, W.N., Henderson, M.C., Granrose, C. S., et al. (1997). Links between race/ethnicity and cultural values as mediated by racial/ethnic identity and moderated by gender. *Journal of Personality and Social Psychology, 72,* 1460–1476.

Giebels, E. & Yang, H. (2009). Preferences for third-party help in workplace conflict: A cross-cultural comparison of Chinese and Dutch employees. *Negotiation and Conflict Management Research, 2*, 344-362.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complimentary justifications for gender inequality. *American Psychologist, 56,* 109-118.

Goldberg, C., Riordan, C. M., & Zhang, L. (2008). Employees' perceptions of their leaders: Is being similar always better? *Group & Organization Management, 33*, 330-355.

Grand, J. A., Ryan, A. M., Schmitt, N., & Hmurovic, J. (2011). How far does stereotype threat reach? The potential detriment of face validity in cognitive ability testing. *Human Performance, 24*, 1-28.

Graves, L. M., & Elsass, P. M. (2005). Sex and sex dissimilarity effects in ongoing teams: Some surprising findings. *Human Relations, 58*, 191-221.

Graves, L. M., & Powell, G. N. (1995). The effect of sex similarity on recruiters' evaluations of actual applicants: A test of the similarity–attraction paradigm. *Personnel Psychology, 48,* 85-98.

Graves, L. M., & Powell, G. N. (1996). Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational and Organizational Psychology, 69*, 243-261.

Harrison, L. A., & Esqueda, C. W. (2001). Race stereotypes and perceptions about Black males involved in interpersonal violence. *Journal of African American Men, 5*, 81-92.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639-683.

Herzog, S., & Oreg, S. (2008). Chivalry and the moderating effect of ambivalent sexism: Individual differences in crime seriousness judgments. *Law & Society Review, 42*, 45-73.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. Organizational Behavior and Human Decision Processes, 81, 226-251.

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin, 28*, 460-471.

Hofstetter, C. H. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16*, 159-188.

Hu, C., Thomas, K. M., & Lance, C. E. (2008). Intentions to initiate mentoring relationships: Understanding the impact of race, proactivity, feelings of deprivation, and relationship roles. *Journal of Social Psychology, 148*, 727-744.

Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science, 37*, 119-147.

Huston, T. L., & Levinger, G. (1978). Interpersonal attraction and relationships. *Annual Review of Psychology, 29*, 115-156.

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly, 37*, 422-447.

Jones, C., & Decotiis, T. A. (1986). Video-assisted selection of hospitality employees. *The Cornell H.R.A. Quarterly, 27,* 68-73.

Judge, T. A., & Livingston, B. A. (2008). Is the gap more than gender? A longitudinal analysis of gender, gender role orientation, and earnings. *Journal of Applied Psychology, 93*, 994-1012.

Kidder, D., & McLean Parks, J. (2001). The good soldier: Who is s(he)? *Journal of Organizational Behavior, 22*, 939-959.

King, E. B., Hebl, M. R., & Beal, D. J. (2009). Conflict and cooperation in diverse workgroups. *Journal of Social Issues, 65*, 261-285.

Kirchmeyer, C. (1995). Demographic similarity to the work group: A longitudinal study of managers at the early career stage. *Journal of Organizational Behavior, 16*, 67-83.

Koch, B. J., & Koch, P. T. (2007). Collectivism, individualism, and outgroup cooperation in a segmented China. *Asia Pacific Journal of Management, 24*, 207-225.

Kozlowski, S. W. J., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology, 77,* 161-167.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70,* 56-65.

Kravitz, D. A., Stone-Romero, E. F., and Ryer, J. A. (1997). Student evaluations of grade appeal procedures: The importance of procedural justice. *Research in Higher Education, 38*, 699-726.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103,* 284-308.

Kuntsman, J. W., & Plant, E. A. (2008). Racing to help: Racial bias in high emergency helping situations. *Journal of Personality and Social Psychology, 95*, 1499-1510.

Landrine, H., Klonoff, E. A., Alcaraz, R., Scott, J., & Wilkins, P. (1995). Multiple variables in discrimination. In B. Lott & D. Maluso (Eds.), *The social psychology of interpersonal discrimination* (pp. 183-224). New York: Guilford Press.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 379-392.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65,* 422-427.

Lau, D. C., Lam, L. W., & Salamon, S. D. (2008). The impact of relational demographics on perceived managerial trustworthiness: Similarity or norms? *The Journal of Social Psychology, 148,* 187-208.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815-852.

Levin, S., Sinclair, S., Veniegas, R. C., & Taylor, P. L. (2002). Perceived discrimination in the context of multiple group memberships. *Psychological Science, 13,* 557-560.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37,* 426-441.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91,* 1181-1188.

Liu, M. (2009). The intrapersonal and interpersonal effects of anger on negotiation strategies: A cross-cultural investigation. *Human Communication Research, 35,* 148-169.

MacDorman, K. F., Coram, J. A., Ho, C-C., & Patel, H. (2010). Gender differences in the impact of presentational factors in human character animation on decisions in ethical dilemmas. *Presence, 19,* 1-17.

MacGeorge, E. L. (2003). Gender differences in attributions and emotions in helping contexts. *Sex Roles, 48,* 175-182.

Martin, B. A. S., Lee, C. K-C., & Yang, F. (2004). The influence of ad model ethnicity and self-referencing on attitudes: Evidence from New Zealand. *Journal of Advertising, 33,* 27-37.

McFarland, L. A., Ryan, A. M., Sacco, J. M., Kriska, S. O. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management. 30,* 435-452.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86,* 730-740.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27,* 415-444.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review, 21,* 402-433.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.

Mount, M. K., Sytsma, M. R., Hazucha, J. F., & Holt, K. E. (1997). Rater–ratee race effects in developmental performance ratings of managers. *Personnel Psychology, 50*, 51-69.

Mumford, T. V., Campion, M. A., & Morgeson, F. P. (2006). Situational judgment in work teams: A team role typology. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 319–343). Mahwah, NJ: Erlbaum.

Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2008). The team role test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology, 93*, 250-267.

Nelson, N. L., & Probst, T. M. (2004). Multiple minority individuals: Multiplying the risk of workplace harassment and discrimination. In J. L. Chin (Ed.), *The psychology of prejudice and discrimination: Ethnicity and multiracial identity* (Vol 2., pp. 193-217). Westport, CT: Praeger Publishers/Greenwood Publishing Group, Inc.

Nguyen, N. T., & McDaniel, M. A. (2003). Response instructions and racial differences in a situational judgment test. *Applied H.R.M. Research, 8*, 33-44.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.

Oppler, S. H., Campbell, J. P., Pulakos, E. D., & Borman, W. C. (1992). Three approaches to the investigation of subgroup bias in performance measurement: Review, results, and conclusions. *Journal of Applied Psychology, 77*, 201-217.

Organ, D. W. (1997). Organizational citizenships behavior: It's construct cleanup time. *Human Performance, 10*, 85-97.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3-72.

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement, 65*, 70-89.

Perkins, L. A., Thomas, K. M., & Taylor, G. A. (2000). Advertising and recruitment: Marketing to minorities. *Psychology & Marketing, 17*, 235-255.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*, 1-16.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26,* 513-563.

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology, 94*, 122-141.

Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770-780.

Reyna, C. (2008). Ian is intelligent but Leshaun is lazy: Antecedents and consequences of attributional stereotypes in the classroom. *European Journal of Psychology and Education, 23*, 439-458.

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology, 85*, 880-887.

Ridgeway, C. L. & Smith-Lovin, K. (1999). The gender system and interaction. *Annual Review of Sociology, 25,* 191-216.

Riordan, C. M., & Shore, L. M. (1997). Demographic diversity and employee attitudes: An empirical examination of relational demography within work units. *Journal of Applied Psychology, 82*, 342-358.

Russell, A. & Owens, L. (1999). Peer estimates of school-aged boys' and girls' aggressionto same- and cross-sex targets. *Social Development, 8,* 364-379.

Sackett, P. R., & DuBois, C. L. Z. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology, 76*, 873-877.

Saulny, S. (2011a, January 30). Black? White? Asian? More young Americans choose all of the above. *The New York Times,* p. A1.

Saulny, S. (2011b, February 10). Counting by race can throw off some numbers. *The New York Times,* p. A1.

Sagar, 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.

Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in helping Whites and Blacks: A meta-analysis. *Personality and Social Psychology Review, 9*, 2-16.

Schmitt, A.P. (1988). Language and cultural characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. *Journal of Educational Measurement, 25*, 1-13.

Shute, R., & Charlton, K. (2006). Anger or compromise? Adolescents' conflict resolution strategies in relation to gender and type of peer relationship. *International Journal of Adolescence and Youth, 13,* 55-69.

Sidanius, J., & Pratto, F. (1999). *Social dominance.* Cambridge: Cambridge University Press.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261-302.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-178.

Thomas, D. A. (1990). The impact of race on managers' experiences of developmental relationships (mentoring and sponsorship): An intra-organizational study. *Journal of Organizational Behavior, 11*, 479-492.

Thomas, K. W., Thomas, G. F., & Schaubhut, N. (2008). Conflict styles of men and women at six organization levels. *International Journal of Conflict Management, 19,* 148-166.

Timberlake, J. M., & Estes, S. B. (2007). Do racial and ethnic stereotypes depend on the sex of target group members? Evidence from a survey-based experiment. *The Sociological Quarterly, 48,* 399-433.

Tjosvold, D. (1998). Cooperative and competitive goal approach to conflict: Accomplishments and challenges. *Applied Psychology: An International Review, 47,* 285-313.

Truxillo, D. M., Bauer, T. N., & Sanchez, R. J. (2001). Multiple dimensions of procedural justice: Longitudinal effects on selection system fairness and test-taking self-efficacy. *International Journal of Selection and Assessment,9,* 336-349.

Tsui, A. S., Egan, T. D., & O'Reilly, C. A. III. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly, 37,* 549-579.

Tsui, A. S., & O'Reilly, C. A. III. (1989). Beyond simple demographic effects: Importance of relational demography in superior-subordinate dyads. *Academy of Management Journal, 32*, 402-423.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogenous and diverse task groups. *Academy of Management Journal, 36*, 590-602.

Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50,* 25-49.

Wesolowski, M. A., & Mossholder, K. W. (1997). Relational demography in supervisor-subordinate dyads: Impact on subordinate job satisfaction, burnout, and perceived procedural justice. *Journal of Organizational Behavior, 18,* 351-362.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Research Management Review, 19,* 188-202.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance, 21*, 291-309.

Whittler, T. E. (1989). Viewers' processing of actors' race and message claims in advertising stimuli. *Psychology and Marketing, 6,* 287-309.

Whittler, T. E. (1991). The effects of actors' race in commercial advertising: Review and extension. *Journal of Advertising, 20,* 54-60.

Whittler, T. E., & DiMeo, J. (1991). Viewers' reactions to racial cues in advertising stimuli. *Journal of Advertising Research, 31,* 37-46.

Williams, R. E., & Qualls, W. J. (1989). Middle-class Black consumers and intensity of ethnic identification. *Psychology and Marketing, 6*, 263-286.

Wilson, L. K., & Gallois, C. (1985). Perceptions of assertive behavior: Sex combination, role appropriateness, and message type. *Sex Roles, 12,* 125-141.

Wilson, K. L., Lizzio, A. J., Whicker, L., Gallois, C., & Price, J. (2003). Effective assertive behavior in the workplace: Responding to unfair criticism. *Journal of Applied Social Psychology, 33,* 362-395.

Zenger, T. R., & Lawrence, B. S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal, 32,* 353-376.