
Are Women Always More Interpersonally Sensitive Than Men? Impact of Goals and Content Domain

Judith A. Hall

Northeastern University

Marianne Schmid Mast

University of Neuchâtel

Two studies examined motivation and content domain as possible influences on sex differences in interpersonal sensitivity. Although much research has found women to excel on tasks measuring interpersonal sensitivity, most of the tasks have measured accuracy in female-relevant domains such as emotion. The present studies measured interpersonal sensitivity, defined as accurate recall of another person, for both female-relevant and male-relevant content domains and also included motivational manipulations intended to influence men and women differently. Study 1 measured accuracy of recalling information in a written vignette about a person, and Study 2 measured accuracy of recalling details about an interaction partner. Both studies supported hypotheses about domain specificity and gender-relevant motivation. However, even for male-stereotypic content and for tasks framed to favor men's motivation to perform well, men's accuracy never exceeded women's.

Keywords: *gender; sex differences; interpersonal sensitivity; motivation; accuracy; appearance; status; dominance*

Interpersonal sensitivity refers to accuracy in processing cues and behaviors in another person. The factors that contribute to accuracy on an interpersonal sensitivity task in general, and to sex differences on such tasks in particular, are not well understood. In general, accuracy could be determined both by preexisting knowledge and by motivation to perform well on the task (Ickes, Gesn, & Graham, 2000; Rosip & Hall, 2004). The relative contributions of these sources of accuracy are unknown at present, and of course the possible role of motivation is not limited to how hard one tries at the moment of testing but also includes the myriad ways in

which motivation may, during a person's lifetime, influence one's habits and one's fund of knowledge about cues and their meanings.

There is a great deal of evidence linking women to skill in interpersonal perception. Women are believed to be more interpersonally sensitive than men, both as a general trait and as a more specific skill in terms of judging the meanings of nonverbal cues (Briton & Hall, 1995; Spence, Helmreich, & Stapp, 1975). They also see themselves as more nonverbally sensitive than men see themselves (Zuckerman & Larrance, 1979). Such beliefs are not unfounded, as research using nonverbal judgment tasks shows that females are more accurate than males on tests of judging the meanings of nonverbal cues conveyed by the face, body, and voice (as shown in meta-analyses by Hall, 1978, 1984; McClure, 2000) as well as on tests of accuracy in recalling another person's nonverbal behavior (Hall, Murphy, & Schmid Mast, 2006) and appearance (Horgan, Schmid Mast, Hall, & Carter, 2004; Schmid Mast & Hall, 2006).

Motivational Influences on Accuracy

Motivational factors in the testing situation have been discussed as possible influences on this sex difference.

Authors' Note: The authors thank Fred Gordon for his theoretical insights and Susanne Christen, Jennifer Publico, Claudia Stadelmann, and especially Jared Majerle for their help in conducting and analyzing these studies. Correspondence should be addressed to Judith A. Hall, Department of Psychology, 125 NI, Northeastern University, Boston, MA 02115; e-mail: j.hall@neu.edu.

PSPB, Vol. 34 No. 1, January 2008 144-155

DOI: 10.1177/0146167207309192

© 2008 by the Society for Personality and Social Psychology, Inc.

Because it is generally evident to the test taker that the task at hand measures ability to process nonverbal information about another person, differential motivation may be stimulated simply by being given such a test, with the result that women try harder (or men try less hard), producing overall superior scores for women. Ickes et al. (2000) proposed further that the sex difference may especially diverge when test takers are given extra reminders about the gendered nature of such skill.

Three experiments support the idea that framing an interpersonal sensitivity test as gender relevant can influence performance. Klein and Hodges (2001) produced better performance among women, but not men, by framing an interpersonal sensitivity task to suggest that the skill is stereotypically female (related to empathy). Horgan and Smith (2006) framed their interpersonal sensitivity test as either stereotypically male (related to military interrogation) or female (related to social work). These manipulations did not improve performance over the level of a control group for either sex, but each sex did worse, compared to the control group, when told the skill was relevant to the opposite sex (military interrogation for women and social work for men). Koenig and Eagly (2005) manipulated participants by emphasizing the female-stereotypic nature of nonverbal cue decoding and produced decrements in the performance of men. All of these studies suggest that motivational effects on an interpersonal sensitivity task can be produced by framing the task as male or female relevant.

In the present studies, we introduced two kinds of manipulations to influence the motivation, and thereby the accuracy, of men and women on interpersonal sensitivity tasks. We call the two kinds of manipulations *goal relevant* and *content relevant*. In both studies, sensitivity was defined as how much participants could remember about a target person who was either described in a vignette (Study 1) or who engaged in a live interaction with the participant (Study 2).

Goal-relevant motivation refers to the implicit or explicit goals aroused by framing or context, or both. In Study 1, the goal-relevant manipulation was intended to influence accuracy especially among women and consisted of telling participants they would be tested on their memory for the person (or not). This prediction is in line with Ickes et al.'s (2000) theorizing because this kind of accuracy would be understood to be a female-relevant skill. In Study 2, the goal-relevant manipulation was intended to influence accuracy especially among men and consisted of placing participants in a competitive (vs. co-action) interaction with a partner while they performed an achievement task. We predicted that being in the competitive condition would have a stronger positive effect on men's than women's accuracy

because it would stimulate the stereotypically male goal of attending to competitors.

In contrast, *content-relevant motivation* refers not to how the task is framed but to the content of the task itself. The content of a sensitivity task may be intrinsically more motivating, or better understood, by one sex than the other. Because the possible effect of content on sex differences in accuracy has rarely been discussed, we devote the next section to this topic.

Female-Relevant Versus Male-Relevant Content

The superior performance of women on tests of interpersonal sensitivity is well documented. In a meta-analysis of 75 studies, Hall (1978) summarized sex differences in decoding the meanings of facial expressions, body movements and postures, and tones of voice. Women scored higher than men in 84% of the studies in which direction could be ascertained, as well as in 96% of the studies achieving statistical significance. In terms of magnitude, the average standardized difference (Cohen's *d*) for studies in which it could be calculated was .40. For a later retrieved collection of 50 studies, the results were nearly identical (Hall, 1984), and studies continue to show this difference (e.g., Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Biehl et al., 1997; Pitterman & Nowicki, 2004; Scherer, Banse, & Wallbott, 2001).

Another large database consists of 133 groups that were given the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), an audiovisual test of accuracy in judging nonverbal cues conveyed by face, body, and voice. In the meta-analyses described in the preceding paragraph, only a few studies using the PONS were included so as not to overrepresent one test. The complete set of 133 PONS studies showed an overall sex difference that was nearly identical in magnitude to that found in Hall's (1978, 1984) meta-analyses.

In a meta-analysis of facial judgment accuracy in children and adolescents, McClure (2000) summarized sex differences for a variety of tests, one of which was developed after the Hall reviews were done. This test, the Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 1994), also showed a significant sex difference favoring females, even though it was designed to minimize such differences. The overall effect size for the DANVA was $d = .18$.

Importantly, the vast majority of studies addressing sex differences in interpersonal sensitivity have used tasks in which the content consists of behavior in affective situations (e.g., asking forgiveness, expressing gratitude) or the expression of emotions (e.g., happiness, sadness, anger). There is no question that affect and

emotion are topics on which women are considered specialists. Women report themselves as experiencing more intense emotions than men report, they are more willing to self-disclose about emotions, and they see themselves, and others see them, as expressing emotions more often and more intensely than men (Brody & Hall, 2000; Cross & Madson, 1997; Gross & John, 1998). Women score higher than men on tests of emotional intelligence (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006). Women's faces are more spontaneously revealing of their emotions than men's faces are, and women are more successful in posing expressions of emotion than men are (Hall, 1984). Fischer and Manstead (2000) found that in all 37 countries in which they gathered data, women rated themselves higher on the nonverbal expression of emotion than men did. Therefore, women's greater accuracy in judging others' states may be specific to the affective domain, where they have special interest and expertise.^{1,2}

Another domain in which women excel is the judgment of personality. Vogt and Colvin (2003) found that women were more accurate in judging a profile of personality traits in target individuals than men were, with an effect size of $d = .80$. Greater accuracy for women in judging specific traits has also been found (Ambady, Hallahan, & Rosenthal, 1995; Lippa & Dietz, 2000; Murphy, Hall, & Colvin, 2003). Understanding others' personalities can be considered female stereotypic because it is congruent with women's orientation toward relating to other people (Cross & Madson, 1997).

Noticing and remembering others' nonverbal behavior is yet another domain in which women score higher than men (Hall, Murphy, et al., 2006). In studies using both videorecorded stimuli and live partners as stimuli, Hall, Murphy, et al. (2006) found that women remembered dynamic cues such as shrugging, smiling, gazing, nodding, licking lips, touching hair, and gesticulating better than men did, with an average Cohen's d of .26. Finally, Horgan et al. (2004) and Schmid Mast and Hall (2006) found, in nine studies, that women remembered others' appearance better than men did, with homogeneous effect sizes and an average Cohen's d of about .50. It can easily be argued that others' physical appearance, especially aspects of clothing and hairstyle, is a female-relevant domain. Indeed, research shows that women have a heightened interest in clothes and appearance compared to men (Kwon, 1997; Schmid Mast & Hall, 2006).

In sum, on tasks having content that is especially relevant to women, women are more accurate than men. Women's superior performance for such content may reflect more accurate knowledge of such cues, women may also have heightened motivation (or men may have reduced motivation) to learn or pay attention to such cues during development, and men and women may

have different motivational responses when presented with such cues on a sensitivity test.

A crucial question raised by the foregoing is whether the female advantage in interpersonal perception would be reversed or reduced for content that is more relevant to men. Several authors have offered this hypothesis (Cross & Madson, 1997; Dovidio & Ellyson, 1982; Haviland & Malatesta, 1981), but it has not received much empirical examination. Dovidio and Ellyson (1982) tested it on a task of attributing degrees of social power to stimulus persons displaying varied proportions of gazing while speaking versus gazing while listening (patterns known to be related to actual social power and therefore considered male relevant). Men and women did not differ in their accuracy on this task, leading Dovidio and Ellyson to speculate that "this result might be explained by the *relatively superior* performance of males in power-related domains compared to affective domains" (p. 111, italics added).

Thus, alternate predictions can be put forth regarding sex differences for a test with male-stereotypic content. If the accuracy difference reverses itself, so that men's interpersonal sensitivity is greater than women's, this would be an *absolute advantage for men*. Alternatively, because any interpersonal judgment task, especially one based on nonverbal cues, may be viewed as female relevant to some extent (Cross & Madson, 1997), men may have a relative, though not an absolute, advantage on tasks having male-stereotypic content, as proposed by Dovidio and Ellyson (1982). According to this *relative advantage for men* hypothesis, the sex difference should be smaller or nonexistent for male content compared to female content, but not reversed in direction. As suggested by Dovidio and Ellyson, the vertical dimension of social relations—differences in power, status, rank, dominance, and related constructs—can be considered a male-relevant domain. Voluminous research shows the connection of the vertical dimension to boys and men, in both fact and stereotype (Maccoby, 1998; Spence et al., 1975). Schmid Mast (2004) showed that men are stereotypically associated with hierarchy and women with egalitarianism, using an implicit association measure. Men have more dominant personalities than women (at least until recently; Twenge, 2001), and they hold stronger expectancies for the emergence of hierarchies than women do (Schmid Mast, 2005).

A few other investigators have also looked at sex differences for male-relevant content. Barnes and Sternberg (1989) asked perceivers to identify who was supervisor and supervisee in photographs and found no significant sex difference (they did not report it in more detail). Schmid Mast, Hall, Murphy, and Colvin (2003) and Schmid Mast and Hall (2004) also measured accuracy on male-content tasks, specifically judging assertiveness

displayed by individuals in videotapes and judging occupational rank from photographs, respectively. Both studies found nonsignificant and negligible sex differences in accuracy.

In keeping with the notion of different sex effects in different domains, the sex difference favoring women is inconsistent on a standard decoding test called the Interpersonal Perception Task (Costanzo & Archer, 1989), which includes items that are both female stereotypic (judging kinship and intimacy) and male stereotypic (judging status and competitive outcomes). Though the norm data for this test showed female advantage ($d_s = .30$ and $.38$ for the long and short forms, respectively; Costanzo & Archer, 1989, 1993), subsequent studies have mainly shown negligible sex differences (Ames & Kammrath, 2004; Aube & Whiffen, 1996; Iizuka, Patterson, & Matchen, 2002; Woods, 1996). These weak results are to be expected if items with female-stereotypic and male-stereotypic content are tallied together in the test's total score.³

Because no study using male-stereotypic content has shown a sex difference favoring men, only the relative advantage for men hypothesis has received any support thus far. However, because there has been little research testing accuracy for male content, it is still possible that an absolute advantage for men could emerge in new studies.

The Present Research

We manipulated both goal-relevant and content-relevant motivation in the present studies to examine their separate and combined effects. Study 1's manipulation of goal relevance was designed to benefit women more than men, whereas Study 2's manipulation of goal relevance was designed to benefit men more than women. In both studies, participants were tested for recall of both female- and male-stereotypic content. Memory for appearance served as the female-stereotypic content area in both studies, with participants being tested for recall of the appearance of either a person described in a written vignette (Study 1) or a live interaction partner (Study 2). Sensitivity to male content was defined in Study 1 as memory for cues indicative of status or dominance in the written vignette, and in Study 2 it was defined as memory for another person's performance on an achievement task. Thus, both studies measured the effects of gender-relevant motivation both with regard to the goals of the task and with regard to the content being tested.

We predicted that, consistent with published findings, women would perform better than men on female-stereotypic content (i.e., appearance). For male-stereotypic content, we predicted either no sex difference, consistent with the relative advantage for men hypothesis, or men would

excel over women, consistent with the absolute advantage for men hypothesis. In addition to these effects, we explored the likelihood that the sex differences for gender-specific content would be greatest in conjunction with gender-specific goal relevance. In Study 1, women's performance should be most favorably influenced by female-stereotypic content plus female-relevant goal, and in Study 2, men's performance should be most favorably influenced by male-stereotypic content plus male-relevant goal.

STUDY 1

Method

Participants

Participants were 236 students (116 males, 120 females) at the University of Zürich majoring in different areas and with an average age of 26 years (range = 18-51). Participants were recruited in classes and were tested either individually or in groups of 5 to 50 people.

Vignette

In the vignette, the sex of the person being described and that person's status/dominance (high vs. low) were crossed, making for four versions, each containing 494 words. There were 26 status/dominance cues (examples: comes from a distinguished/not distinguished family, completed college degree/dropped out of college, chooses expensive/inexpensive restaurants, likes exclusive red wine/a cool can of beer, drives BMW/Opel, plays golf/volleyball, likes to make the decisions at work/lets others make the decisions, thinks it is important/not important to assert him/herself, and likes/does not like to take on leadership position). None of the status/dominance cues pertained to appearance. Hereafter, the status/dominance cues are referred to simply as status cues.

There were 42 appearance cues (examples: athletic build; brown hair; short hair; silvery frame on glasses; rather angular face; high cheekbones; straight nose; scar above eyebrow; wears hair combed out of the face; rather fine hair; hair often falls back across forehead; wears wedding band; wears a wristwatch; wears blue, brown, or black coats; wears jeans). There were also 43 "other" cues that were concerned with neither status nor appearance (examples: married for 2 years, has no children, has no sisters, has no brothers, parents are dead, occasionally goes back to the town where grew up, most friends live in Zürich, has relatively extended circle of friends, enjoys books, particularly fascinated with thrillers, loves cats, has no pets, had bicycle accident as child, met spouse 4 years ago). All 111 cues could apply equally to either sex.

TABLE 1: Recall of Appearance, Status, and Other Cues, Study 1 (Frequencies)

Type of Accuracy and Goal Condition	Men	Women	Effect Size (<i>d</i>)
Recall of appearance cues			
Control	12.98	15.37 ⁺	.34
Female goal	13.19	19.63 ^{***}	.87
Total	13.08	17.50 ^{***}	.61
Recall of status cues			
Control	9.52	11.02 [*]	.45
Female goal	8.79	11.08 ^{***}	.83
Total	9.15	11.05 ^{***}	.61
Recall of other cues			
Control	13.78	17.58 ^{***}	.80
Female goal	14.60	19.57 ^{***}	.87
Total	14.19	18.58 ^{***}	.80

NOTE: The significance levels indicate male–female differences. Positive effect sizes reflect higher accuracy by women than men. ⁺ $p < .10$. ^{*} $p < .05$. ^{***} $p < .001$.

Female-Goal Manipulation

In the control condition, participants read that they would be forming an impression of the person in the vignette, with no mention of the need to remember details or the fact that they would be tested on their recall. In the female-goal condition, they read that they would be asked to remember as much as they could about the person. Participants were not told what kind of information would be in the vignette.

Procedure

Participants were randomly assigned to one of the eight conditions (Female Goal/Control × Male/Female Target × High/Low Status). Participants were given 3 min to read the vignette, after which the experimenter collected it and gave them a sheet of blank paper on which to report as many details about the person as they remembered. Participants were given a maximum of 10 min to complete this task, though most participants needed less time.

Scoring of Recall

Two trained coders counted for each participant the number of status items remembered (range = 2–24, $M = 10.12$, $SD = 3.34$), the number of appearance items remembered (range = 1–39, $M = 15.34$, $SD = 7.93$), and the number of other items remembered (range = 3–37, $M = 16.45$, $SD = 6.06$). Incorrect responses were so infrequent that they were not systematically counted. Intercoder reliability between the two coders based on 60 participants was $r = .97$ for status items, $r = .99$ for appearance items, and $r = .97$ for other items. Each coder scored recall for half of the participants.

Results

We conducted separate four-way, between-subject ANOVAs on the number of appearance cues recalled and the number of status cues recalled. The independent variables were participant sex (male/female), target sex (male/female), target status (high/low), and goal manipulation (control/female goal). We discuss only effects involving participant sex and we report effect sizes for main and simple effects only.⁴ Table 1 shows the means for these analyses.

For accuracy of recalling appearance, participant sex produced a highly significant main effect, $F(1, 220) = 20.50$, $p < .001$ (Table 1, top panel). There was also a Participant Sex × Goal Manipulation interaction, $F(1, 220) = 4.34$, $p < .05$. The means revealed that women's advantage over men was more pronounced in the female-goal condition than in the control condition. Another way to look at this effect is to note that the female-goal manipulation had only a minimal effect on men but it had a pronounced effect on women.

The same ANOVA for number of status cues recalled produced only a participant sex main effect, $F(1, 220) = 20.71$, $p < .005$ (Table 1, middle panel). It showed that women recalled more status cues than men did, and the effect was of identical magnitude as in the preceding result. Because there was no Participant Sex × Goal Manipulation interaction ($F < 1$), one can conclude that arousing a female-relevant goal in women did not produce any additional benefits when male-stereotypic content was being recalled.

The results from the same ANOVA using recall of other cues as the dependent variable showed a participant sex main effect, $F(1, 220) = 35.38$, $p < .001$, which again showed that women performed better than men (Table 1, bottom panel). As with recall of status, there was no Participant Sex × Goal Manipulation interaction ($F < 1$). Thus, for both male content (status) and other content, and unlike the situation for female content (appearance), framing the task as an accuracy task brought no extra benefits to women.

Because there were unequal numbers of male- and female-stereotypic cues, analysis of raw frequencies did not allow for a direct comparison between these domains. To do this, we calculated percentage accuracy scores by dividing the number of each person's recalled cues of each type by the total number of available cues of that type. For example, because there were 42 appearance cues, a person who remembered 15 would obtain a percentage accuracy score of 36% in this domain. Focusing on results that were nonredundant with those already presented, we found that women's percentage accuracy did not differ between appearance and status ($M = 42\%$ for both), whereas men's percentage

accuracy was lower for appearance ($M = 31\%$) than for status ($M = 35\%$). The latter difference was significant in a separate repeated measures ANOVA for men, $F(1, 114) = 6.97, p < .01$. Thus, women were equally good in both domains, whereas men were more accurate for status than for appearance.

Because women excelled on all three kinds of cues (female stereotypic, male stereotypic, and other), we undertook a final analysis that examined how men's and women's recall was distributed between appearance and status cues in a relative sense. This is different from the preceding analysis because that analysis was concerned with absolute accuracy (how many cues were recalled out of the total possible), whereas this analysis is concerned with an individual's pattern of responding and can therefore be considered an idiographic approach. For each participant we calculated proportion scores defined as the participant's proportions of appearance and status cues recalled out of his or her total number of cues recalled. Using each participant's total recall as the denominator for these proportions meant that the scores were independent of each person's overall level of accuracy and therefore independent of the strong overall sex main effect. The presence of other cues meant that the status and appearance proportion scores were not negatively collinear with each other, though they were negatively correlated, $r(234) = -.62, p < .001$.

We conducted a five-way ANOVA on these proportion scores with the same independent variables as used previously (participant sex, target sex, target status, and goal manipulation) and with content (appearance or status) as a repeated measures factor. No participant sex main effect was expected for this analysis because a sex difference in the pattern of recall would cancel out when collapsed over the two kinds of content; indeed, the participant sex main effect was $F < 1$. However, there was a significant Participant Sex \times Content interaction, $F(1, 220) = 6.47, p = .01$, showing that women recalled proportionately more appearance cues than men did, and men recalled proportionately more status cues than women did (Table 2, lines labeled "Total").

Furthermore, the goal manipulation had a moderating effect, as reflected in the Goal Manipulation \times Participant Sex \times Content interaction, $F(1, 220) = 3.32, p = .07$ (Table 2). Men's behavior was unaffected by the female-goal manipulation, but women's behavior changed dramatically, such that women shifted their proportional attention (as indicated by recall) toward appearance and away from status when in the female-goal condition. Separate ANOVAs on men and women confirmed this interpretation: For men, the Goal Manipulation \times Content interaction was $F < 1$, whereas for women it was $F(1, 112) = 12.68, p < .001$. Pairwise tests revealed that women's increase in proportional attention to appearance and decrease in proportional

TABLE 2: Recall of Appearance and Status, Study 1 (Relative Proportions)

Type of Accuracy and Goal Condition	Men	Women	Effect Size (d)
Recall of appearance cues			
Control	.34	.34	.02
Female goal	.33	.39**	.59
Total	.34	.36+	.26
Recall of status cues			
Control	.27	.26	.22
Female goal	.26	.22**	.58
Total	.27	.24**	.38

NOTE: The significance levels indicate male-female differences. Positive effect sizes reflect higher accuracy by women than men. * $p < .10$. ** $p < .01$.

attention to status were both significant ($p < .01$; means shown in Table 2).

Discussion

Men and women were asked to remember details about a person from a written narrative that contained information about status (male-stereotypic content) and personal appearance (female-stereotypic content), as well as other characteristics. As a motivational trigger that we thought would positively affect women, we manipulated the salience of accurate recall by telling some participants before they read the narrative that their accuracy would be tested. Ickes et al. (2000) proposed that women's often-reported superiority in interpersonal sensitivity may stem not so much from what they know but from their motivation to perform well on a task they consider to be stereotypic for women, especially if the gender-stereotypic nature of the skill is made salient. Ickes et al. found, for a collection of studies using the empathic accuracy paradigm (see note 1), that the sex difference favoring women was much more apparent when participants were reminded of the gender-stereotypic nature of the task (by being required to evaluate their own accuracy after every judgment) than when no such requirement was in place. Subtle manipulations of gender relevance such as those used by Ickes et al. and by us in Study 1 may be optimal for influencing behavior, as more blatant manipulations may result in paradoxical effects (e.g., efforts to defy rather than conform to the stereotype for one's sex). Other studies that have found motivational effects based on manipulations of gender relevance have also used relatively subtle manipulations (Horgan & Smith, 2006; Klein & Hodges, 2001).

However, simply being asked to form an overall impression of a person (control condition) produced substantially better recall in women than in men regardless of the content of the cues. Therefore, the act of

drawing an impression of a person by itself reflected gender-linked motivation or habits of information processing, or both, consistent with the notion that any kind of interpersonal information processing may be stereotypically female. But in addition, we expected that the gender relevance of the content (status vs. appearance) and the female-goal manipulation (to remember as much as possible about the person) would additionally stimulate gender-relevant motives that would further influence accuracy. Consistent with this expectation, the female-goal manipulation produced a dramatic gain in accuracy selectively for women on the female-stereotypic task of remembering the person's appearance. The goal manipulation had no effect on men, and it had no effect on women for either male-stereotypic or other content.

The analysis of the within-participant pattern of recall, conducted on the proportions of the two types of cues recalled out of a participant's total recall, shed further light on the sex differences. Women and men distributed their recall differently, with women remembering proportionately more appearance cues than men did and men remembering proportionately more status cues than women did. This was especially evident in the female-goal condition. The female-goal condition did not influence men's proportion scores, whereas that condition caused women to recall proportionately more appearance cues and proportionately fewer status cues.

The relative advantage for men hypothesis was not supported in terms of the overall number of male- and female-stereotypic cues remembered by the sexes, as women's advantage was of the same magnitude for both types of cues and was also marked for the other cues (i.e., neither male nor female stereotypic). However, the analysis of proportions suggested that men did give preferential attention to the male-stereotypic cues over the female-stereotypic cues in a relative sense. Furthermore, the analysis of percentage accuracy showed that women's performance was the same for both appearance and status cues, whereas men's performance was worse for appearance than for status cues. Though not directly supporting the relative advantage for men hypothesis, this result indicates that the content domain did influence men's accuracy.

In sum, Study 1 showed that sex differences in interpersonal sensitivity, defined as recalling written information about a person, are related to both gender-relevant content and gender-relevant goals. In Study 2, we moved from a vignette to a live interaction paradigm to replicate and extend these findings.

STUDY 2

In Study 2, we again compared male and female recall accuracy in male-stereotypic and female-stereotypic

content domains, in conjunction with a gender-relevant goal manipulation. Participants interacted with a partner without being told they would be asked to recall information about the partner after the interaction. The male-stereotypic task was to recall the partner's performance on the anagram-solving task they had each been performing (a recall task adapted from Saenz, 1994). We reasoned that monitoring a partner's performance on a skill-based task has implications for status ranking (who is doing better than whom) and is therefore male stereotypic in content. The female-stereotypic recall task was again to recall the partner's personal appearance. Gender-relevant goal motivation was manipulated by randomly assigning participants to a co-action or competition condition while solving the anagrams. We surmised that competition would motivate men, more than women, to attend to the partner's performance, which in turn would favorably influence men's accuracy of performance recall. However, competition should not influence men's accuracy of remembering the partner's appearance because the partner's appearance was not relevant to their goals. Whereas in Study 1 the goal manipulation (to think of the task as assessing accuracy in processing cues about another person) was intended to increase women's recall accuracy, in Study 2 the goal manipulation (to compete with the partner) was expected to increase men's recall accuracy.

Method

Participants

Participants were 106 students (54 males, 52 females) recruited from the Psychology Department Participant Pool at Northeastern University as partial fulfillment of their introductory course requirement. According to self-report, 76.4% were Caucasian, 3.8% were African American, 10.4% were Asian American, 5.7% were Hispanic, and 3.7% were Other or unreported ethnicity; their mean age was 18.65 (range = 17-22). One male and one female served as experimenters.

Task

Participants had to solve five-letter anagrams, each of which was printed on a separate piece of paper. One hundred anagrams were constructed to range in difficulty from easy to hard, following Tresselt and Mayzner's (1966) guidelines.

Male-Goal Manipulation

All participants were placed in dyads. In the control (co-action) condition, participants were instructed to work individually to solve the anagrams and not to talk during the task. The male-goal (competition) condition

was the same, but participants were also told that they should complete as many anagrams as possible and that the person who completed more than the partner would be awarded \$5.00 at the end of the session.

Procedure

Participants were randomly assigned to same- or mixed-sex dyads and to control (co-action) or male-goal (competition) conditions. Once in the laboratory, participants were seated across from each other and given the informed consent. Participants were then told they would solve anagrams of varying difficulty for 15 min. Each participant had an open plastic tray on the table, into which he or she was instructed to place the anagrams face down as they were completed. The anagrams were numbered but were stacked in a random order so that participants could not keep track of how many anagrams had been completed simply by remembering the number of the last anagram worked on.

While the participants were completing the anagrams, the experimenter monitored the participants to make sure directions were followed. The experimenter started a stopwatch when the participants began the task and gave reminders when there were 10 min, 5 min, and 1 min left. Following the anagrams, the experimenter explained to participants that they would be completing some questionnaires in separate rooms.⁵ These questionnaires contained the recall variables that would be scored for accuracy. First, participants were asked to estimate as accurately as possible how many anagrams they had completed and how many their partner had completed. Following this, participants were asked to write down everything they could remember about their partner's appearance, under the categories of hair, shirt or top, pants or skirt, shoes, and general aspects such as nationality, makeup, jewelry, and physical features. Following this, participants wrote down a description of their own appearance using the same categories. After the questionnaires (which included demographic questions) were completed, participants were brought together and debriefed. In the competition condition, the prize was awarded.

Scoring of Recall

Performance recall. Participants estimated they had completed, on average, 24.52 anagrams ($SD = 14.28$) and their partner had completed 26.05 ($SD = 11.19$). Overall, participants solved an average of 30.98 anagrams ($SD = 14.14$). Performance recall accuracy was calculated by comparing estimated performance with actual performance in terms of the absolute proportion of over- and underestimation. Larger scores indicated more error (i.e., the estimate was farther from the actual

TABLE 3: Recall of Appearance and Performance, Study 2

Type of Accuracy and Goal Condition	Men	Women	Effect Size (d)
Better recall of partner's appearance (high = better recall)			
Control	5.95	7.33 ⁺	.54
Male goal	6.43	7.54	.45
Total	6.19	7.44 [*]	.49
Error in recall of own task performance (high = worse recall)			
Control	0.27	0.32	.16
Male goal	0.29	0.30	.08
Total	0.28	0.31	.12
Error in recall of partner's task performance (high = worse recall)			
Control	0.37	0.25 ⁺	.56
Male goal	0.24	0.28	-.18
Total	0.30	0.27	.18

NOTE: The significance levels indicate male-female differences. Positive effect sizes reflect higher accuracy by women than men. ⁺ $p < .10$. ^{*} $p < .05$.

number). This method was used to score self-error and partner performance error (self-error: $M = .30$, $SD = .24$; partner error: $M = .29$, $SD = .20$).

Appearance recall. Appearance recall was calculated by comparing participants' descriptions of their partner's appearance with the partner's self-description. Experimenter observations were also made for most of the participants and these were used. For each description match, 1 point was awarded to the participant's appearance recall accuracy score. For further details on this method, see Horgan et al. (2004). A high number of points indicated high accuracy in appearance recall ($M = 6.78$, $SD = 2.56$, range = 2-13).

Results

Separate between-subject ANOVAs were conducted for the three accuracy measures, in which participant sex (male/female), partner sex (male/female), and goal manipulation (control/male goal) were the independent variables.

For recall of the partner's appearance, the participant sex effect was significant, $F(1, 96) = 5.85$, $p < .05$, with women being more accurate than men. Participant sex did not interact with goal manipulation ($F < 1$), meaning that the sex difference for recalling appearance was similar in both conditions.⁶ Table 3 shows the means for this analysis.

For error in recall of own anagram performance, we expected no effects and this was confirmed ($ps > .23$). For error in recall of partner performance, there was no

overall participant sex effect ($F < 1$), consistent with the relative advantage for men hypothesis. There was a marginally significant Participant Sex \times Goal Manipulation interaction, $F(1, 97) = 3.34, p = .07$. The means, shown in Table 3, indicate that women's performance was the same in both conditions but men's performance improved in the male-goal (competition) condition—that is, fewer errors were made. In the control condition, women's performance advantage over men was close to significant by conventional standards, $F(1, 47) = 3.68, p = .06$, but in the male-goal condition, men made slightly fewer errors than women, though the difference was negligible ($F < 1$). Separate ANOVAs for men and women showed that for men, performance was significantly better in the male-goal condition than in the control condition, $F(1, 50) = 4.21, p < .05$, whereas for women this effect was negligible ($F < 1$).

Accuracy in recalling the partner's appearance and error in recalling the partner's performance were uncorrelated, $r(101) = .02$.

Discussion

In Study 2, we assessed accuracy of recalling a partner in a female-stereotypic domain (appearance) and a male-stereotypic domain (task performance) while manipulating goal motivation by having participants work on the task in parallel (control condition) or in competition for a prize (male-goal condition). Consistent with previous research (Horgan et al., 2004; Schmid Mast & Hall, 2006), women were more accurate than men in recalling the partner's appearance (see also note 6). Women were also more accurate (made fewer errors) than men in recalling the partner's performance in the control condition, suggesting a generally greater interpersonal focus among women, but men's accuracy came up to the level of women's in the male-goal condition. Thus, men were motivated by competition to be more accurate than they would otherwise be, whereas the motivational manipulation had no effect on women. Analysis of error in remembering one's own performance was not related to any of the experimental factors, which gives reassurance that the results do not reflect simply an overall higher level of vigilance or better memory.

This study, like Study 1, supports the hypothesis that motivational factors play a role in sex differences in interpersonal sensitivity. A male-relevant goal (competition) increased men's performance on the male-stereotypic sensitivity task beyond the level they otherwise would have displayed, but it did not influence their performance on the female-stereotypic sensitivity task. The relative advantage for men hypothesis was thus supported in this condition because the gap between men

and women was smaller for the male-stereotypic task (remembering the partner's performance) than for the female-stereotypic task (remembering the partner's appearance). However, in the control condition the relative advantage for men hypothesis was not supported because the sex difference was of the same magnitude for both tasks (see Table 3).

GENERAL DISCUSSION

In two studies, accuracy of recalling information about another person was investigated in relation to gender-relevant content and gender-relevant goals. Recall of physical appearance was the female-stereotypic content in both studies, and recall of social status cues and recall of partner performance on a skill task were the male-stereotypic content in Studies 1 and 2, respectively. Two alternate hypotheses about the effect of content on sex differences in accuracy were tested, one proposing an absolute advantage for men over women in male-stereotypic domains and the other proposing a relative advantage for men in male-stereotypic domains. The latter hypothesis acknowledges that any task involving interpersonal sensitivity, or recall of personal attributes or behaviors, may be considered stereotypically female and therefore can be expected to show a female advantage. However, the relative size of the female advantage could vary with the gender stereotypicality of the domain such that females' advantage is smaller in male-stereotypic domains and with gender-specific goals such that each sex gains from goals that are especially relevant to their own sex.

We found no support for the absolute advantage for men hypothesis, as men were not significantly more accurate than women in any comparison, whereas women were often more accurate than men. There was limited support for the relative advantage for men hypothesis. In Study 1, women's advantage was of the same magnitude in both male- and female-stereotypic domains, and in Study 2 women's advantage was smaller in the male-stereotypic than female-stereotypic domain only in the male-goal condition.

We might speculate that because solving anagrams involves verbal skill, which is a female-stereotypic trait, Study 2 did not go as far as it could have in creating a male-stereotypic recall task. Thus, though the task of recalling a partner's performance on a skill task may be male stereotypic, it might have been even more so had the activity itself been more blatantly stereotypic for men—for example, solving math problems, playing a male-stereotypic sport, or performing a male-stereotypic mechanical task. The inclusion of such tasks in future studies will provide a further test of the domain specificity hypothesis.

There was a moderating effect of the gender-relevant goals. In Study 1, women's accuracy in the female-relevant domain was increased in the female-goal condition, and in Study 2 men's accuracy in the male-relevant domain was increased in the male-goal condition. In that study, framing the task in terms of competition added sufficient male relevance to motivate men to be more accurate than they would otherwise be on the male-stereotypic task, but they were still not more accurate than women. Thus, we obtained insight into factors influencing accuracy of interpersonal perception for men and women, but we did not fully explain the sex differences.

An important question, though one that is beyond the scope of the present research, is the extent to which sex differences in interpersonal sensitivity are based on knowledge versus motivation. Though researchers have typically assumed that the differences are based on knowledge of cues and their meanings, and though women score higher than men on a paper-and-pencil test of knowledge of nonverbal cue usages and meanings (Rosip & Hall, 2004), the sex differences seen on actual sensitivity tasks could stem from a mix of knowledge and motivation. Furthermore, over one's lifetime, knowledge and motivation are intertwined because motivation to attend to cues probably leads to enhanced knowledge about those cues, which leads to enhanced interest and attention to those cues, and so on.

Although one might argue that defining interpersonal sensitivity in terms of recall of cues (as opposed to inferring meanings of cues) helps simplify the picture because recall might be considered to depend only on motivation (effort) and not preexisting knowledge, we think that domain knowledge may still be relevant to recall accuracy. Greater knowledge of a domain facilitates the encoding and retrieval of information in that domain (Bransford, Brown, & Cocking, 1999). Certainly, future research must include a broad range of interpersonal sensitivity tasks because one should not generalize from a limited methodology. Especially, effects of motivation on tests that draw more on preexisting knowledge (i.e., where obtaining the correct answer depends on drawing accurate connections between cues and their meanings) must be examined more than they have been. This is important because increased motivation is not necessarily synonymous with improved performance; this may be especially true for social inference tasks that draw on preexisting knowledge more than recall tasks do.

There has been some theorizing about the relation between gender and interpersonal sensitivity as it relates to a person's dominance or social power, where it is argued that women are more interpersonally sensitive than men because women's lower status and dominance motivate them to be accurate (Henley, 1977). However,

though authors have speculated that low status and dominance motivate higher accuracy (Keltner, Gruenfeld, & Anderson, 2003) and that high status and dominance motivate inaccuracy (Fiske & Morling, 1996), evidence for a negative association between status and dominance and interpersonal sensitivity is rare (Galinsky, Magee, Inesi, & Gruenfeld, 2006), with most research finding either that the association is positive (e.g., meta-analysis by Hall, Halberstadt, & O'Brien, 1997; Schmid Mast & Jonas, 2007) or that the association is negative but due to other factors (e.g., Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Snodgrass, Hecht, & Ploutz-Snyder, 1998). Though it is not impossible that women's lower status and dominance will be shown to have a causal role in their interpersonal accuracy, at present the more fruitful avenue seems to be in terms of the effect of gender stereotypes and gender-relevant motivational factors on knowledge and performance.

Understanding of the roots of sex differences in interpersonal sensitivity is progressing but it has a long way to go. The difficulty of this undertaking is not surprising considering that sex differences are the product of many forces impinging on individuals over their lifetimes as well as at the time of testing. The present research contributes to our understanding by examining the gender stereotypicality of the domain and the gender relevance of participants' goals as moderators of the sex differences.

NOTES

1. Another method for testing interpersonal judgment accuracy, not included in the cited meta-analyses, is called the empathic accuracy paradigm (Ickes, 2001; Ickes, Gesn, & Graham, 2000; Klein & Hodges, 2001; Thomas & Fletcher, 2003). In this paradigm, dyad members guess each other's thoughts and feelings while watching a video replay of their interaction (or, alternatively, while watching a video of strangers in conversation). Accuracy is scored by comparing the perceiver's guesses about the target's thoughts and feelings with the target person's statements about his or her thoughts and feelings. Studies using this method generally show less of a female advantage than found using more standard testing paradigms (though still not favoring men). This might seem to contradict the notion that women excel in judging others' affective cues. However, the empathic accuracy paradigm may not be primarily about judging affective states because perceivers are asked to judge thoughts as well as feelings; indeed, accuracy for this paradigm depends much more on the verbal statements by target persons than on their nonverbal cues, which can be expected to carry much of the affective information (Gesn & Ickes, 1999; Hall & Schmid Mast, 2007). Therefore, it is not clear to what extent research using the empathic accuracy paradigm contradicts the general conclusion that women excel at judging affective cues.

2. One might predict that sex differences vary with the male or female relevance of the specific emotions that are being judged. However, the predictions that one should make are far from clear. As an example, stereotype might suggest that anger is a male-relevant emotion; on the other hand, self-report studies do not find that men report experiencing more anger than women do (Brody & Hall, 2000). In fact, one could argue that judging anger is female relevant because of women's risk of victimization. A meta-analysis found a relatively large female decoding advantage for decoding anger (Bauer,

Kulkarni, & McGowan, 1997). Establishing predictions for judgments of fear is equally problematic. One might predict that men, as the intimidators who produce fear in others, should be especially good at judging that emotion. On the other hand, one might predict superiority for women because they report feeling more fear than men do (Brody & Hall, 2000). Bauer et al. (1997) found that this emotion showed the biggest female advantage of all the emotions they compared. The examples of anger and fear indicate that it is premature to reach any conclusions about male versus female domain specificity in regard to the decoding of specific emotions.

3. The Interpersonal Perception Task (IPT; Costanzo & Archer, 1989) contains items that test lie-detection accuracy, a skill we argue is ambiguous in its gender connotations. Consistent with the IPT's often negligible sex differences, Aamodt and Custer's (2006) meta-analysis of 53 lie-detection studies found no sex difference ($d = .03$).

4. Interactions that included both participant sex and target sex would have been potentially interesting, but they were negligible in both studies (Study 1: $.13 < p < .99$; Study 2: $.23 < p < .89$), with the exception of one 4-way interaction in Study 1 that was not interpretable.

5. Participants who decided to sit in the chair closest to the exit were arbitrarily labeled Participant B and the other participant was labeled Participant A. Correlations between Participants A and B were nonsignificant for all of the accuracy scores: for appearance recall, $r(50) = .23, p = .09$; for recall of other's performance, $r(50) = .02$; and for recall of own performance, $r(51) = .04$.

6. A condition in which participants were instructed to cooperate in solving the anagrams was also included. Because participants did not work separately, accuracy of remembering the partner's performance could not be calculated; therefore, this condition is not discussed in this article. However, participants' accuracy of remembering the partner's appearance was calculated for this condition. Women were significantly ($p < .05$) more accurate than men ($d = .61$), just as in the control and male-goal conditions.

REFERENCES

- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner, 15*, 6-11.
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518-529.
- Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*, 187-209.
- Aube, J., & Whiffen, V. E. (1996). Depressive styles and social acuity: Further evidence for distinct interpersonal correlates of dependency and self-criticism. *Communication Research, 23*, 407-424.
- Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence, 13*, 263-287.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241-251.
- Bauer, L. M., Kulkarni, M., & McGowan, S. (1997). *Gender differences in nonverbal decoding of emotion: A meta-analysis*. Unpublished manuscript, University at Albany, State University of New York.
- Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3-21.
- Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*, 780-795.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles, 32*, 79-90.
- Brody, L. R., & Hall, J. A. (2000). Gender, emotion, and expression. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2d ed., pp. 338-349). New York: Guilford.
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225-245.
- Costanzo, M., & Archer, D. (1993). *The Interpersonal Perception Task-15 (IPT-15): A guide for researchers and teachers*. Berkeley: University of California Extension.
- Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin, 122*, 5-37.
- Dovidio, J. F., & Ellyson, S. L. (1982). Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly, 45*, 106-113.
- Fischer, A. H., & Manstead, A. S. R. (2000). The relation between gender and emotions in different cultures. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 71-94). Paris: Cambridge University Press.
- Fiske, S. T., & Morling, B. (1996). Stereotyping as a function of personal control motives and capacity constraints: The odd couple of power and anxiety. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Vol. 3. The interpersonal context* (pp. 322-346). New York: Guilford.
- Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science, 17*, 1068-1074.
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology, 77*, 746-761.
- Gross, J. J., & John, O. P. (1998). Mapping the domain of expressivity: Multimethod evidence for a hierarchical model. *Journal of Personality and Social Psychology, 74*, 170-191.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*, 845-857.
- Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.
- Hall, J. A., Halberstadt, A. G., & O'Brien, C. E. (1997). "Subordination" and nonverbal sensitivity: A study and synthesis of findings based on trait measures. *Sex Roles, 37*, 295-317.
- Hall, J. A., Murphy, N. A., & Schmid Mast, M. (2006). Recall of nonverbal cues: Exploring a new definition of interpersonal sensitivity. *Journal of Nonverbal Behavior, 30*, 141-155.
- Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology, 42*, 18-27.
- Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion, 7*, 438-446.
- Haviland, J. J., & Malatesta, C. Z. (1981). The development of sex differences in nonverbal signals: Fallacies, facts, and fantasies. In C. Mayo & N. M. Henley (Eds.), *Gender and nonverbal behavior* (pp. 183-208). New York: Springer-Verlag.
- Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication*. Upper Saddle River, NJ: Prentice Hall.
- Horgan, T. G., Schmid Mast, M., Hall, J. A., & Carter, J. D. (2004). Gender differences in memory for the appearance of others. *Personality and Social Psychology Bulletin, 30*, 185-196.
- Horgan, T. G., & Smith, J. L. (2006). Interpersonal reasons for interpersonal perceptions: Gender-incongruent purpose goals and nonverbal judgment accuracy. *Journal of Nonverbal Behavior, 30*, 127-140.
- Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219-241). Mahwah, NJ: Lawrence Erlbaum.

- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95-109.
- Iizuka, Y., Patterson, M. L., & Matchen, J. C. (2002). Accuracy and confidence on the Interpersonal Perception Task: A Japanese-American comparison. *Journal of Nonverbal Behavior*, 26, 159-174.
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, 110, 265-284.
- Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720-730.
- Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, 52, 489-496.
- Kwon, Y. (1997). Sex, sex-role, facial attractiveness, social self-esteem and interest in clothing. *Perceptual & Motor Skills*, 84, 899-907.
- Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior*, 24, 25-43.
- Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126, 424-453.
- Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality*, 71, 466-493.
- Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, 18, 9-35.
- Pitteman, H., & Nowicki, S., Jr. (2004). A test of the ability to identify emotion in human standing and sitting postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Test (DANVA2-POS). *Genetic, Social, and General Psychology Monographs*, 130, 146-162.
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.
- Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28, 267-286.
- Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition*, 12, 61-74.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76-92.
- Schmid Mast, M. (2004). Men are hierarchical, women are egalitarian: An implicit gender stereotype. *Swiss Journal of Psychology*, 63, 107-111.
- Schmid Mast, M. (2005). Interpersonal hierarchy expectation: Introduction of a new construct. *Journal of Personality Assessment*, 84, 287-295.
- Schmid Mast, M., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, 28, 145-165.
- Schmid Mast, M., & Hall, J. A. (2006). Women's advantage at remembering others' appearance: A systematic look at the why and when of a gender difference. *Personality and Social Psychology Bulletin*, 32, 353-364.
- Schmid Mast, M., Hall, J. A., Murphy, N. A., & Colvin, C. R. (2003). Judging assertiveness. *Acta Universitatis*, 2, 731-744.
- Schmid Mast, M., & Jonas, K. (2007). *It's power that makes you interpersonally sensitive!* Unpublished manuscript.
- Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238-249.
- Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, 32, 29-39.
- Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, 85, 1079-1094.
- Tresselt, M. E., & Mayzner, M. S. (1966). Normative solution times for a sample of 134 solution words and 378 associated anagrams. *Psychonomic Monograph Supplements*, 1, 293-298.
- Twenge, J. M. (2001). Changes in women's assertiveness in response to status and roles: A cross-temporal meta-analysis, 1931-1993. *Journal of Personality and Social Psychology*, 81, 133-145.
- Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, 71, 268-295.
- Woods, E. (1996). Associations of nonverbal decoding ability with indices of person-centered communicative ability. *Communication Reports*, 9, 13-22.
- Zuckerman, M., & Larrance, D. T. (1979). Individual differences in perceived encoding and decoding abilities. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences* (pp. 171-203). Cambridge, MA: Oelgeschlager, Gunn & Hain.

Received November 27, 2006

Revision accepted June 27, 2007