Wesleyan University The Honors College

Does the Scientist's Sex Matter in the Laboratory?

The Effects of an Experimenter's Sex in Psychology Experiments

by

Brittany Morse

Class of 2008

A thesis submitted to the
faculty of Wesleyan University
in partial fulfillment of the requirements for the
Degree of Bachelor of Arts
with Departmental Honors in Psychology

Middletown, Connecticut April, 2008

Acknowledgments

I would like to extend my sincerest gratitude to Professor Jill Morawski for advising me on this thesis. Her constant guidance, direction, help, encouragement, and support have been invaluable. I appreciate the opportunities, resources, and knowledge with which she has provided me, along with the reliable assistance and constructive feedback that she has offered throughout this process.

I would also like to thank Jeffrey Decew, Katie Feldman, John Gottfried, Adam Greenson, Daniel Lachman, Andrew Meyer, Catherine Porfert, Ezra Silk, and Lauren Westling for their contributions and assistance.

Table of Contents

Introduction 4

Literature Review 7

Study 1: Meta-analysis 19

Method 19

Results 21

Discussion 37

Study 2: Experiment 48

Method 48

Results 57

Discussion 70

Appendices 83

Appendix A 83

Appendix B 93

Appendix C 104

Appendix D 106

Appendix E 111

References 112

Coded studies 116

Introduction

The objectivity of scientific research, with its plethora of methodological procedures, is rarely questioned. After controlling variables that might obscure data, a study's reliability and validity are often taken for granted. With such assurance, psychological researchers commonly ignore the fact that experimenters (just like the participants with whom they interact) are human beings with individual identities—identities comprising race, personal history, socioeconomic status, educational background, appearance, affect, personality, and sex. These and other experimenter characteristics are invariably present, and could influence participants and confound data in ways that are difficult to detect, analyze, and correct. This thesis examines the effects of one of these characteristics: experimenter sex.

Examination of experimenter sex, arguably one of the experimenter's most salient traits, exposes potential biases that may permeate the conduct and interpretation of research. Through interrogation of the presence (or absence) of experimenter sex bias, along with its specific insinuations, comes a deeper understanding of the ways in which certain aspects of the experimenter-participant relationship—and certain aspects of the analytic process—influence research outcomes. This investigation thus should be of inherent interest to psychological experimenters, and to those outside the discipline insofar as experimenter sex effects may have evocative implications for sexed interactions beyond the laboratory.

The literature review section of the thesis provides chronological syntheses of reviews, meta-analyses, and observations that provide a framework for conducting two studies on experimenter bias and sex bias in psychology research. The increasing

acknowledgment of the experimenter as a subjective individual, the historical role of women in psychology, and the causes and consequences of first-, second-, and third-wave feminism provide the context for examining experimenter sex.

Study 1 is a qualitative meta-analysis of 144 empirical studies that consider experimenter sex effects, and is part of a larger study being conducted by Jill Morawski at Wesleyan University. A detailed survey of the studies reveals the occurrence and findings of studies on experimenter sex over time. It shows trends in researchers' explanations of findings and in studies' methodological limitations, considered with respect to contextual factors that characterize specific periods of time. Study 2 is a modified replication of a frequently cited empirical study (Piacente, 1974) on participants' perceptions of male and female experimenters. The variables hypothesized to influence participants' judgments include experimenter sex and an "experimenter condition" (competent versus incompetent). Other independent variables include participant sex and administrator sex. Analyses of independent samples t-tests reveal main effects, interaction effects, and second-order interaction effects of these variables. The findings are compared with those of the original study, their causes are theorized, and their implications—hypothetical and concrete—are outlined.

The insights provided by the present studies increase appreciation for the complexity and vicissitudes that mark the frequently ignored intricacies of the unavoidably sexed features of research. Psychological experimenters concerned with the internal and external validity of empirical studies can benefit from a greater awareness of the results of studies on experimenter sex effects, and from a deeper

knowledge of their significance. More broadly, a general understanding of the invariable sexing of hierarchical interactions and judgments may shed light on the dynamics of power relations between the sexes at large.

Literature Review

A unique virtue of experimentation lies in the isolation of specific variables, allowing experimenters to systematically manipulate and measure potentially causal factors. By comparing experiments that isolate the same variable, psychological researchers can develop and/or substantiate theories about human behaviors and cognitive processes. A disadvantage of human psychological experimentation lies in the sometimes dubious reliability of such studies. It is logistically impossible to consider, control, and test for every potential influencing variable; thus experimenters must wonder: is a study measuring the issues that it is intended to measure, and nothing else? What factors surreptitiously influence a study's participants, unbeknownst to the experimenter? For instance, the data collected in an experiment relying on responses to questionnaires might be affected by the experimenter's mode of participant solicitation (telephone versus airmail versus advertisement versus in-person recruitment). An experiment on children's behaviors might be influenced by the fact that the experiment is administered in a laboratory rather than in the field.

In the 1950s and 1960s, researchers began to acknowledge that participants may respond to factors (such as method of solicitation or setting of the study) distinct from the intended independent variables. Accordingly, attention to potential confounding variables (such as biased sampling and lack of equivalence between laboratory and field studies) grew. One factor researchers brought to the table for critical analyses was experimenter bias.

A leading researcher in the study of experimental conditions, Rosenthal (1966), noted, "Some of the complexity of man as we know it from his model, the

research subject, resides not in the subject himself but rather in the particular experimenter and in the interaction between subject and experimenter" (p. vii). A contemporary of Rosenthal, Barber (1968), reported that an experimenter's subjectivities often influence research findings. Rosenthal, along with Barber, pioneered acknowledgment of experimenter bias, and soon researchers began identifying specific experimenter characteristics and designing studies that empirically explored these characteristics. Experimenter sex/gender is one of the regarded variables. About a half century later, Morawski (2006) commented on the problems caused by psychology's historical failure to consider further how the "[s]tandpoint—both of the observer and the observed—and experience…are inseparable from the process of knowledge seeking and the very knowledge ultimately produced" (p. 5).

Experimenters used the term "sex" almost exclusively until the 1980s, when "gender" came into use. Since the 1980s, experimenters have used "sex" and "gender" in comparable frequency and often interchangeably, despite the suggestion by researchers and theorists that "sex" refer to one's biology and that "gender" refer to one's personal and social identification (Unger, 1979). Because "sex" is the more historically prevalent term (used in the majority of studies reviewed in the present meta-analysis), "sex" will henceforth be the operative term in this text.

Experimenter sex has been proven a crucial variable to study as research has shown that it often has significant effects on participants. If experimenter sex effects persist through time, then experimenters should consider them when designing studies and when interpreting studies' results. From a logical standpoint, consistent

empirical evidence of experimenter sex bias would suggest researchers' obligation to vary experimenter sex in *every* ensuing study. From a theoretical standpoint, evidence of experimenter sex bias should compel researchers to explore *why* such effects take place—what psychological processes (in the experimenters as well as in the participants) produce such effects.

Research has indicated that experimenter sex may affect a study's results in four important ways. Experimenter sex may influence: 1) participants' behaviors or responses to the experimental stimuli; 2) participants' perceptions of the experimenter; 3) the experimenter's interpretations of his/her findings; and 4) the experimenter's reporting of his/her findings. Empirical research has focused almost solely on the first two ways: the effects of experimenter sex on participants' behaviors and perceptions. Meta-analyses and literature reviews, on the other hand, have focused largely on the latter two ways: the effects of sex on an experimenter's interpretations and reporting of findings.

High interest in experimenter sex effects coincided with two events: concern with methodological issues (namely experimenter bias) and feminism. As noted, concern with confounding factors in human laboratory experimentation grew in the 1950s and 1960s. The insurgence of second-wave feminism in the 1960s heightened attention to psychology's largely androcentric methodologies.

Sex bias through the feminist movements of the 20th century

In the late 19th century, psychology emerged as a discipline almost exclusively conducted by, for, and about men (Shields, 1975). Commonplace

procedures included utilizing solely male experimenters and participants, sex-biased stimuli and procedures, and androcentric modes of explaining research results (Rossiter, 1982). These male biases have roots in the discipline's structure. Few women were members of the American Psychological Association (founded in 1892) in its early years, as G. Stanley Hall, the association's founder, viewed intellectually oriented women as "functionally castrated" (Scarborough, 1987, p. 4). Clearly, at the end of the 19th century, women were not expected to play a significant role or to be seriously considered in psychology research.

The American Psychological Association's founding coincided with first-wave feminism, a movement advocating equal rights between men and women, specifically with regard to the right to vote. While feminist leaders entered the realms of social and political activism, feminist leaders emerged in the sciences as well. Early scientific reformers such as Leta Hollingworth and Helen Thompson Woolley challenged experimental psychology's methodologies in the 1910s (Russo, 1983; Rosenberg, 1982). Later, when the American Psychological Association's Emergency Committee in Psychology (formed in preparation for war assistance) was found to have no female representation, a group of women in New York organized the National Council of Women Psychologists (Russo, 1983). Thus, feminist criticisms of the androcentric aspects of psychology were raised in the science's early years, and continued intermittently until the 1960s.

Critical attention to the role of women in psychology increased dramatically through the 1960s and peaked in 1970s with second-wave feminism. After World War II, career-related opportunities for women declined, along with affirmations of

women's independence and agency, inspiring a wave of feminism focused on encouraging women to broaden their spheres of influence beyond the domestic. In turn, feminists in the sciences noted and sought to deconstruct the androcentrism that dominated scientific disciplines. In 1968, Naomi Weisstein commented on "the uselessness of present psychology…with regard to women" (Hare-Mustin, 1990). Astin (1973) reiterated the pervasion of sex-based discriminatory practices in all areas of the discipline and Gray (1977) called attention to the androcentrism of psychology textbooks. Astin's concerns in particular, with a focus on equalizing hiring practices and questioning stereotypic notions of masculinity and femininity, directly reflected the social and political pursuits of second-wave feminism.

Socially and politically, first- and second-wave feminism culminated in a few important legislative milestones, most notably the passages of the 19th amendment in 1920 (granting women the right to vote) and Title IX in 1972 (granting equal funding for men's and women's education-related programs). In the wake of attaining such goals, current third-wave feminists have attended to the "micro-politics" of sex arrangements; such as the morality of sex work and the deconstruction of first- and second-wave feminism's "essentialist" implication that upper-middle-class, heterosexual, white women represented female identity at large. Psychological researchers of this period fittingly have sought to distinguish between the unconcealed, deliberate sexism to which the first two waves of feminism responded and the more subtle, idiosyncratic sexism to which third-wave feminists respond.

Glick (1997) differentiates hostile and "benevolent" sexism: while hostile sexism is blatantly misogynistic, benevolent sexism endorses paternalism and

idealization of women. Glick argues that hostile and benevolent sexism alike serve to justify and maintain patriarchal systems. Through administration of the validated Ambivalent Sexism Inventory, Glick found that both men and women in the United States exhibit higher rates of benevolent sexism than hostile sexism (Glick, 2001). Abrams (2003), Sibley (2004), Barreto (2005), and Dardenne (2007) are a few of the many researchers to conduct follow-up studies, uniformly reinforcing Glick's assertion of the problems inherent in benevolent sexism, despite its innocent pretenses. Benevolent sexism is more persistent than hostile sexism likely because it is harder than hostile sexism to prevent, detect, and erase.

Criticisms of experimenter and sex bias in psychology research

In the 1970s and 1980s, at the crux of second-wave feminist activism, sexist methodologies in psychology research came under particularly sharp scrutiny. Harris (1971) noted psychologists' frequent neglect to consider experimenter sex in designing, analyzing, and reporting studies. Harris affirmed the need for such analysis, reporting that experimenter sex significantly influences research on schizophrenia, psychological testing, and sexual attitudes and behaviors. Diamond (1976) cited bias in society, biased use of test content, and biased use of test results as three principal sources of experimenter sex bias in measurement of psychological test results. Eizner (1977) commented that review articles often over-value tasks at which males excel and under-value tasks at which females excel, and that journals more readily publish reports of sex differences than reports of sex similarities, exaggerating the degree to which sex differences occur. Mead (1978) reinforced Harris' (1971) and Diamond's

(1976) imperative reports by arguing that research on sex-specific behaviors should always be conducted by both male and female experimenters in order to correct for prejudice, bias, and myopia. Grady (1981) elaborated that sex bias could taint methodologies and obscure results through several means; specifically those involving the selection of topics, hypotheses, methods, variables, and the analyses and interpretations of data. McHugh (1986) agreed that sexism may be inadvertently present through a researcher's confidence in traditional (largely androcentric) methods of research: through selection of research topics and participants that may produce results more pertinent to males than to females, through bias in explanatory systems particularly with regard to terminologies and variable measurements employed, and through inappropriate generalization of results. Denmark (1988) echoed Eizner (1977), Grady (1981), and McHugh (1986) with a plea to researchers to avoid sexism in their studies by considering sex stereotypes in question formulation, sample selection and variable labeling, and by taking care not to inaccurately magnify gender differences or fabricate erroneous conclusions about data.

Shortly therafter, Condor (1991) commented that ethical guidelines in psychology research have had little effect in preventing the use of sexist research techniques, indicating a pervasive lack of concern for feminist issues. Sherif (1994) further decried the discipline's unrelenting theoretical and research bias, wondering if there had simply not been enough women in psychology to oust sex bias, or if there was something intrinsically problematic about psychology's assumptions and working practices that would take more than female presence to exonerate. Unger

(2007) reiterated the biasing tendency, noted by Eizner (1977) and Denmark (1988), to inaccurately magnify sex differences, primarily by testing for and reporting differences more frequently than similarities between men and women.

The many criticisms of experimenter bias and sex bias in psychology research especially in the 1970s and 1980s logically coincide with a surge in empirical studies on experimenter sex effects. The surge was precipitated synthetically by factors such as the insistence of second-wave feminist interests, advancements in psychology such as the American Psychological Association's forming a division to focus on women (Mednick, 1975), and the publication of Rosenthal's (1966) review on experimenter sex effects in psychology research.

Reviews of empirical research on experimenter and sex bias

Rosenthal's (1966) review of empirical analyses of experimenter sex effects reported, "[M]ale and female experimenters sometimes obtain significantly different data from their subjects" (p. 42). Rosenthal described the complexity and contradictions of the studies' findings, and thus intimated the importance of a more comprehensive meta-analysis of studies on experimenter sex effects. No meta-analyses specifically on experimenter sex bias in psychology research published more recently than Rosenthal's (1966) review can be located. However, two meta-analyses with narrower foci (Rumenik, 1977; Eagly, 1978) and a meta-analysis with a broader focus (Gannon, 1992) are relevant.

Rumenik (1977) reviewed studies on experimenter sex effects in "behavioral" research. Rumenik's review revealed that while conclusions were hard to draw as a

consequence of common methodological inadequacies in the research, experimenter sex was frequently a potent, if fluctuating, variable. Eagly (1978) reviewed studies on sex differences in influenceability, echoing Rosenthal's description of complex and contradictory findings. Eagly reported, "[T]he differences between studies with 1970s publication dates and those with pre-1970s publication dates [and] the contemporary growth of the women's movement" indicate a "concomitant shift in sex role attitudes" (p. 102), however, "biases in role assignment…accord men more power than women" (p. 107). Eagly thus acknowledged feminist progress while simultaneously bringing attention to the persistence of unequal power relations between the sexes. Gannon (1992) reviewed 4,952 studies published from 1970 to 1990 on sex bias (as opposed to *experimenter* sex bias) in psychology research. She commented that studies using data collected with only male participants to make claims about females were becoming increasingly rare, and that sexist language in studies' reports had been virtually eliminated. Gannon's comprehensive review suggested that, through the virtual elimination of sexist language, through a decrease in inappropriate generalizations of results, and through evolving editorial policy and research practices, "sexism in psychology has been diminished but not eliminated" since 1970. Taken together, Gannon's review and Glick's (1997) research on hostile versus benevolent sexism give the impression that, while progress away from psychology's androcentrism and sexism has been achieved, there is certainly more progress to be made.

Critiques of problematic methods in psychology research

While neglecting consideration of experimenter sex is a methodological limitation in and of itself, researchers manipulating the variable often fall prey to the deficits caused by other methodological constraints. In analyzing studies on experimenter sex, one must remain cognizant of the problematic methods in the research. In fact, methodological inadequacies are so ubiquitous in research on experimenter sex effects that Rumenik (1977), as noted, was reluctant to draw conclusions about the findings of studies in her review.

Rumenik's criticisms focused largely on the problems implicit in the use of only one male and one female experimenter: "It is obvious (or should be) that any differences found between one male and one female experimenter may be due to numerous non-sex-related variables on which the two may differ" (p. 874), such as appearance, affect, race, age, attire, or status.

Another methodological shortcoming that might produce data reflecting participants' responses to experimenter traits aside from sex is the use of only undergraduate students as experimenters. Psychology research relies widely on participant samples comprising solely undergraduate students (introductory psychology students in particular) because, ethically and logistically, it is often not feasible to obtain a more diverse and representative sample. Although this reliance has been periodically criticized, little has changed. Sears (1986) held that undergraduate students are markedly unique in a number of ways, limiting the generalizability of studies using only undergraduate student samples. This lack of generalizability was reinforced by a review (Gordon, 1986) of 32 studies, in which

samples of undergraduate students and samples of other people were tested under identical conditions and the results compared, reporting that significant between-group differences were often revealed. Jung (1969) mentioned other concerns (pertaining to recruitment methods, justification of required participation, and confidentiality) arising from the fact that, as Smart (1966) decried, undergraduate students are vastly overrepresented as research participants. The issues raised by Sears, Gordon, and Jung are relevant to our knowledge of experimenters as well as participants. By virtue of the specific experimenter demographic characteristics (particularly with regard to age, socioeconomic status, marital status, and educational background) implicit in studies using only undergraduate students as experimenters, such studies may fail to produce generalizable results. Furthermore, there may be a considerable discrepancy between researchers' recruitment and treatment of undergraduate student experimenters and researchers' recruitment and treatment of other experimenters, potentially obscuring studies' findings.

Another methodological oversight relating to experimenter characteristics is that researchers often neglect to consider and report experimenter demographics beyond sex. Silverman (1974) rebuked researchers for their overwhelming failure to specify and systematically vary the demographic characteristics of experimenters in psychology studies. Disregard for description and consideration of experimenter demographics was further criticized by Guthrie (1976, 1997) and Morawski (1997) in reviews of studies ignoring the potentially confounding effects of experimenter race. Other researchers, such as Carringer (1974), Dutton (1974) and Barnes (1985), have likewise acknowledged the potentially confounding and interacting effects of

experimenter characteristics by varying traits such as sex, race, socioeconomic status, liberalism, attractiveness, and attire in empirical research.

Study 1: Meta-analysis

Amidst the proliferation of literature explaining, criticizing, and analyzing experimenter bias and sex bias, and scrutinizing the accompanying methodological flaws, experimental studies manipulating experimenter sex have employed an array of dependent variables. This meta-analysis considers the main and interaction effects reported in empirical studies, researchers' explanations of results, and problematic methods in experiment design.

Method

The meta-analysis examines the consideration of experimenter sex effects in empirical (largely experimental) psychology research. The analysis updates and expands upon Rosenthal's (1966) review of experimenter sex effects by discerning patterns of research findings over time (from the 1920s to the present), dependent variables measured, explanations of empirical results, and problematic methods. The analysis aims to increase our understanding of the ways in which psychology research is connected with and/or affected by social changes in sex roles; specifically, in relation to the feminist movements of the 20th century.

The hypotheses stated that: 1) the number of studies on experimenter sex effects in psychology research would surge during the 1970s and taper to the levels of previous years thereafter; and 2) experimenter sex effects would be revealed in over half the empirical studies, including those published recently. There are two reasons for these hypotheses. First, the 1970s were marked by a mounting presence of published literature by and about women in science—an upsurge that faded with the

quieting of the second-wave feminist voice. Second, because sex is an undeniably potent variable in the calculations of everyday interactions, it seems logical that sex would impact interactions in the laboratory.

Using PsycINFO as a database, I searched for empirical psychology studies that examined experimenter sex effects in research. I then examined the works cited in each of these studies and obtained relevant studies that were not already located. After limiting the search to exclude non-empirical reviews and studies published in languages other than English, there remained 277 studies, published from 1925 to the present.

A coding template was created to record core features of the studies: reasons for study, hypotheses, experimenters used, subjects/participants used, experiment design, results, explanations of results, implications, recommendations, noted limitations, and problematic methodologies. Over half the studies were randomly selected from each decade and gathered using Ovid, Illiad (an interlibrary loan service), and the two libraries at Wesleyan University, Olin Library and Exley Science Library. Two coders established inter-rater reliability by comparing the completed coding templates of a sample of studies until coding concurrence was obtained. Together, 144 studies were coded, comprising a representative sample of the 277 studies collected. (See Appendix A for examples of studies coded with the template from each decade from the 1950s to the present.)

Using Microsoft Excel, coding results were transferred onto spreadsheets to facilitate recognition of patterns. The frequency of empirical research on experimenter sex effects over time first was examined; then, the resulting frequency patterns across

time were considered in terms of how they related to the contemporaneous literature on women in psychology, feminism, and experimenter bias.

The studies analyzed focus on an array of dependent variables, such as the choices children participants make while playing, participants' rates of being "conditioned", and participants' helpfulness, talkativeness, and compliance. The studies were divided by key dependent variables measured. The occurrence of each dependent variable was recorded, the most prevalent dependent variables determined, and frequencies charted over time.

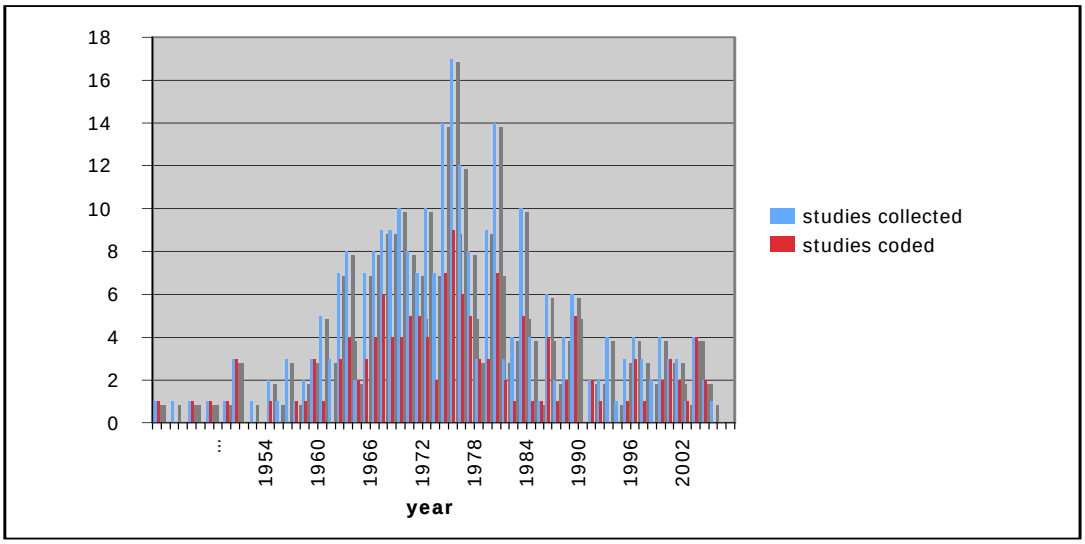
In examining the explanations of results provided by researchers, several models of explanation appeared repeatedly. The most recurring explanatory paradigms were identified, analyzed, and the frequencies charted over time. The frequencies and temporalities of specific problematic methodologies in coded studies were also determined.

Results

Occurrence of studies

Graph 1 shows the number of studies published across time that were collected and coded.

Graph 1

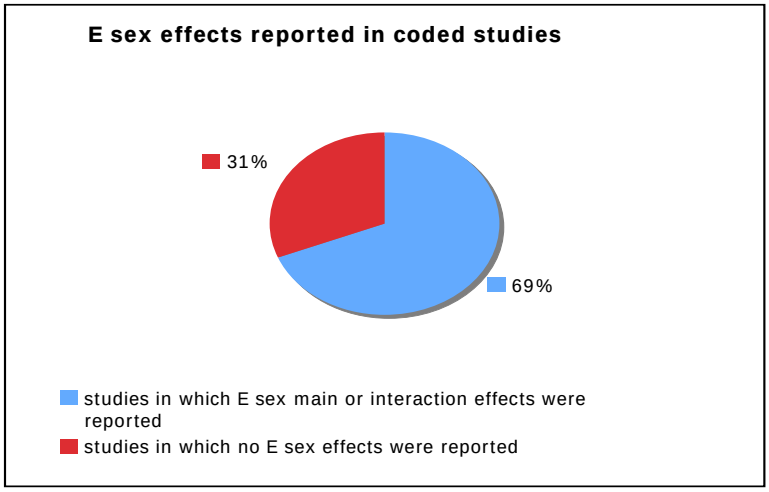


As the graph illustrates, there were fewer than four empirical studies per year on experimenter sex effects until 1962. Out of the 277 empirical studies collected, 126 of them (nearly half) were published between 1967 and 1979. About a third of *those* studies were published between 1976 and 1978, when interest in the topic peaked. These data partly confirm the first hypothesis: studies on experimenter sex effects surged in the 1970s. However, the second part of the hypothesis predicted that the occurrence of post-second wave feminist research on experimenter sex would match the occurrence of research in previous years. Clearly, this is not the case. Although the occurrence of studies decreased dramatically after the second-wave feminist tide, the number of studies per year has remained relatively constant since then. Furthermore, the number of studies per year is notably higher than the number of studies per year prior to second-wave feminist influence.

Experimenter sex effects

The present meta-analysis follows up on Rosenthal's (1966), Rumenik's (1977), and Eagly's (1978) reports of inconsistent results among studies on experimenter sex effects. The present analysis indicates that male and female experimenters often do obtain significantly different data from their participants. Just over two-thirds of studies report, as illustrated by Graph 2, that experimenter (E) sex has either main or interaction effects; confirming the second hypothesis, which posited that experimenter sex effects would be revealed in over half the coded studies, including those published recently.

Graph 2



The effects found through the present analysis are of multiple kinds: 1) sex of experimenter main effects; 2) sex of experimenter-sex of participant interaction effects; 3) sex of experimenter-other variable interaction effects; and 4) sex of participant main effects. The findings are illustrated by Graph 3 and by Table 1.

Graph 3 shows the number, and Table 1 shows the percentage, of coded studies per decade that report sex of experimenter (SOE) main and interaction effects and sex of participant (SOP) main effects.

Graph 3

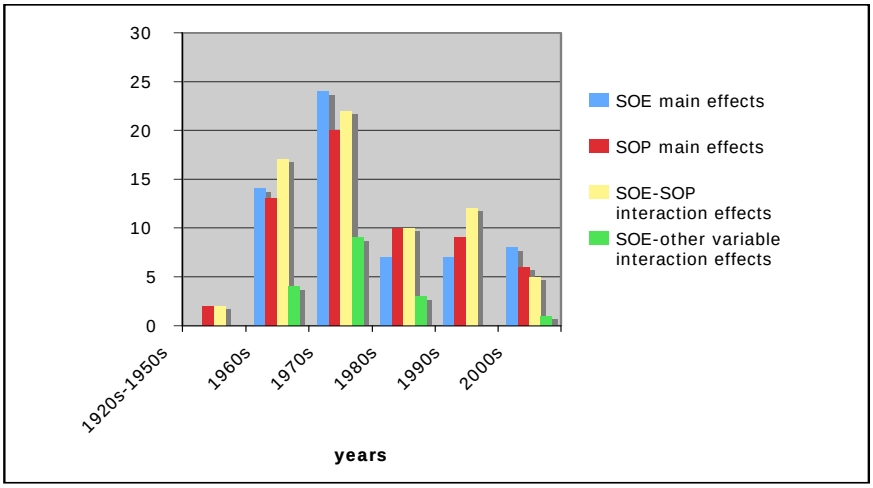


Table 1

Years	# of coded studies	SOE main effect(s) reported	SOP main effect(s) present reported	SOE-SOP interac-tion effect(s) reported	SOE-other interac-tion effect(s) reported	At least one SOE main or SOE interaction effect reported
1920-1950s	9	0%	44%	56%	0%	56%
1960s	27	52%	48%	63%	15%	81%
1970s	51	47%	39%	43%	18%	78%
1980s	28	25%	36%	36%	11%	54%
1990s	15	47%	60%	47%	0%	64%
2000s	14	54%	43%	36%	7%	79%

As the graph and table show, sex of experimenter-sex of participant interaction effects are most common, appearing more frequently than sex of experimenter main effects, sex of experimenter-other variable interaction effects, or sex of participant main effects. Experimenter sex main effects are reported in about half the coded studies except those published before 1960 or during the 1980s. Reports of sex of experimenter-other variable interaction effects are particularly sparse.

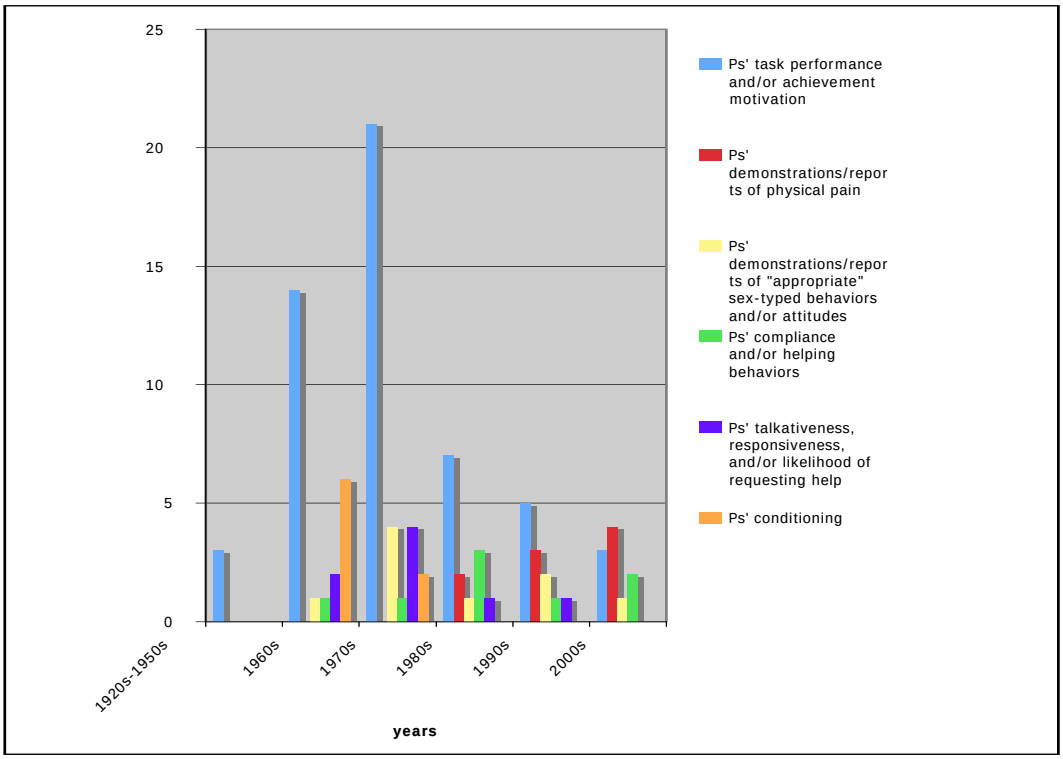
Although patterns in types of effects are revealed, these patterns are meaningless without context. Experimenter sex main effects are revealed in about half the coded studies, but what does this mean? Specifically, *how* do participants respond differently to male versus female experimenters? In order to answer these questions, we must consider each study in relation to other studies examining the same or similar dependent variables.

Experimenter sex effects with respect to specific dependent variables

The most recurrent dependent variables pertained to participants' behaviors. They included: task performance and/or achievement motivation; likelihood of reporting physical pain; demonstrations of "sex-appropriate" behaviors; compliance and helpfulness; talkativeness, responsiveness, and willingness to request help; and levels of conditioning. Graph 4 shows the frequency over time of the six most common dependent variables: participants' (Ps') task performance and/or achievement motivation was by far the most prevalent, measured in over a third of the

coded studies; followed by participants' demonstrations/reports of physical pain and their demonstrations/reports of "sex-appropriate" behaviors.

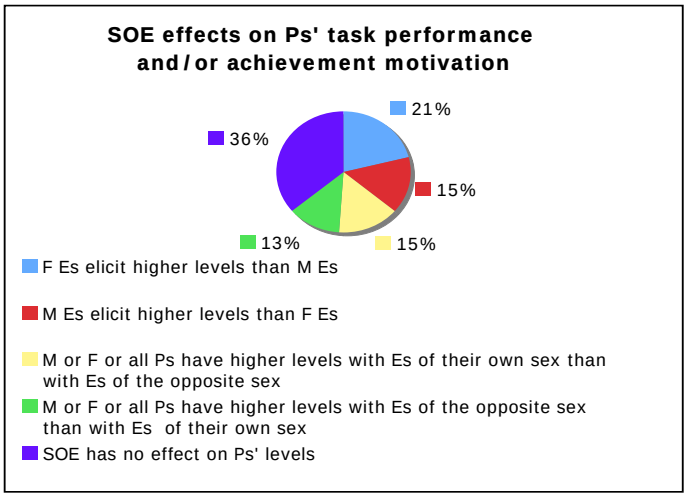
Graph 4



Some consistencies with regard to main and interaction effects emerge when studies on specific dependent variables are considered independently. Graphs 5-10 show the experimenter sex effects reported in studies on specific dependent variables.

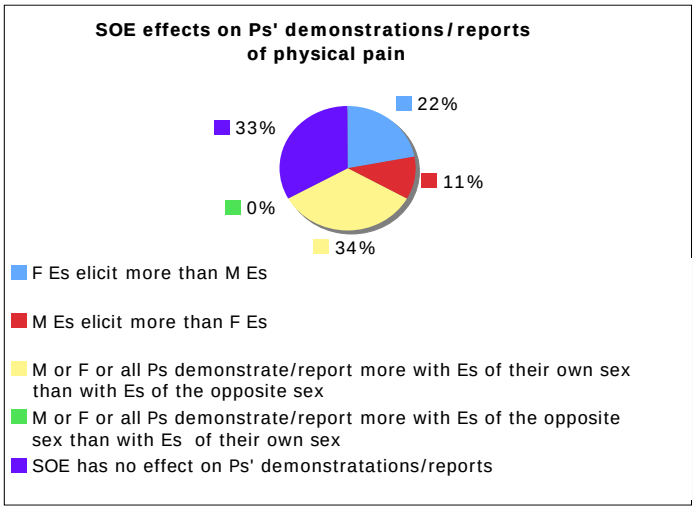
Graph 5 shows that participants' task performance and/or achievement motivation are unaffected by experimenter sex—male (M) or female (F)—in one third of studies, while the four measured kinds of main and interaction effects occur in comparable frequencies, with a slight indication that task performance and achievement motivation may be higher with female than with male experimenters.

Graph 5



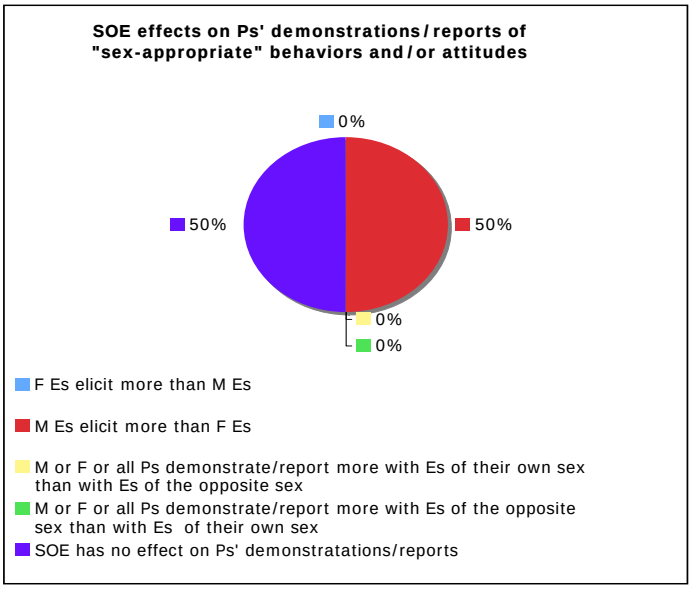
Graph 6 reveals that participants are more likely to report physical pain to female experimenters than to male experimenters and to experimenters of their own sex than to experimenters of the opposite sex. About one third of studies, however, report that participants' reports of pain are unaffected by experimenter sex.

Graph 6



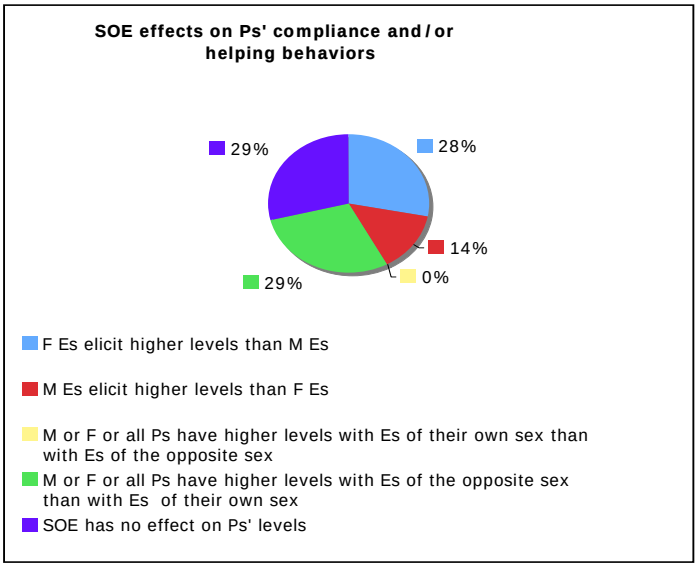
Graph 7 reveals another main effect: participants are more likely to behave in a “sex-appropriate” manner with male than with female experimenters. Half of studies, however, report that displays of “sex-appropriate” behavior are unaffected by experimenter sex.

Graph 7



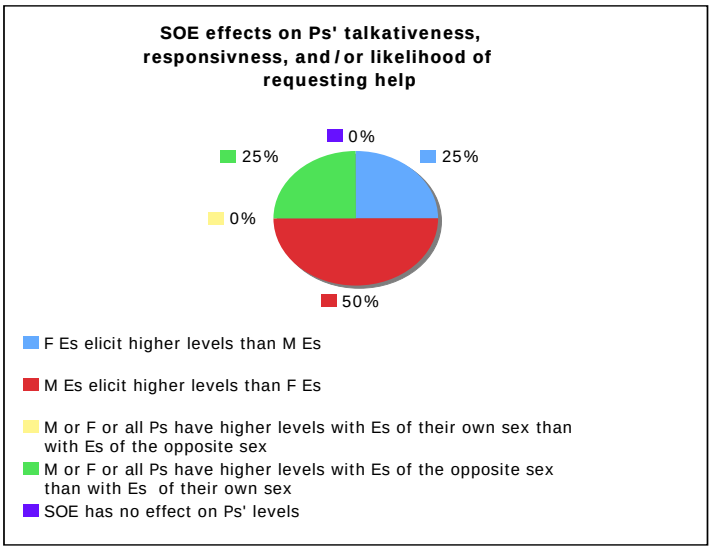
Graph 8 shows that participants are more compliant and helpful with female than with male experimenters, and with experimenters of the opposite sex than with experimenters of their own sex. Nearly one third of studies, however, report that compliance and/or helping behavior are unaffected by experimenter sex.

Graph 8



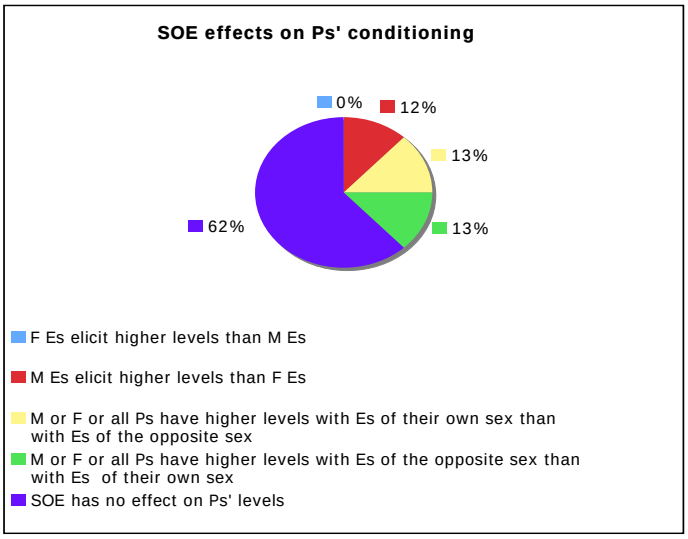
Graph 9 shows that participants are more talkative, responsive, and likely to request help with male than with female experimenters and with experimenters of the opposite sex than with experimenters of their own sex.

Graph 9



Graph 10 shows that research on participants' conditioning reveals a sex of experimenter main effect: participants tend to condition more easily with male than with female experimenters. The majority of studies, however, report that conditioning is unaffected by experimenter sex.

Graph 10



Reported explanations of findings

While some explanations of results provided by researchers are specific to the particular study described, a number of explanatory models appear across many studies. Examination of reported explanations reveals repetition of certain explanatory models. For instance, the likening of experimenter-participant interactions to parent-child interactions appears repeatedly. If a specific explanatory model appears in at least 10% of studies in a given decade (the decades preceding 1960 are grouped together), then that paradigm is considered part of a trend.

Experimenter-participant and parent-child interactions (1960s-1970s)

Researchers in the 1960s and 1970s often explain experimenter sex effects by pointing to the likeness between experimenter-participant interactions and parent-child interactions (Stevenson, 1961; Stevenson, 1964; Burton, 1966; Odom, 1966; Datta, 1968; Quereshi, 1968; Doll, 1971; Stein, 1971; Harney, 1972; Garrett, 1974; Galbraith, 1976; McMahan, 1776). Researchers suggest that participants regard experimenters in a similar manner as they regard parental figures. Specifically, many researchers refer to Freudian conceptions of the Oedipal complex, positing that very young children display greater attachment to the mother than to the father, and that somewhat older children display greater attachment to the parent of the opposite sex than to the parent of the child's own sex. For example, in a study on factors influencing the resistance to temptation, Burton (1966) reports that four-year-olds cheat more in a beanbag-throwing task with experimenters of their own sex than with experimenters of the opposite sex. Burton explains that the child of four is experiencing increasing libidinal attachment toward the opposite-sex parent, motivating the child to behave well (for instance, by not cheating in a beanbag-throwing contest) in the presence of an opposite-sex experimenter.

Sex-typing of the participants' task (1970s)

Researchers in the 1970s consider the possibility that participants respond to the "masculinity/femininity" of the task at hand (O'Sullivan, 1973; Rikli, 1976; Halperin, 1977; White, 1978; Baucom, 1979). With this explanatory model, researchers propose that male and female participants' responses to male and female

experimenters depends on the nature of the task, particularly its association with masculinity or femininity, involved in the experiment. For example, O' Sullivan (1973) reports that male participants improve more than female participants in a modeling task that involves approaching a harmless snake. The researcher suggests that males improve more than females because there exists a stigma against a fear of snakes in males, while there is no such stigma for females. In other words, the researcher believes that the "masculinity" of the task of approaching snakes influenced the participants.

Experimenters' personal characteristics (1940s-1980s)

Between the 1940s and 1980s, researchers increasingly point to the personal characteristics (aside from sex) of individual experimenters as a likely confounding factor in studies on experimenter sex effects (Remmers, 1940; Alden, 1951; Archer, 1961; Sarason, 1965; Gold, 1969; Johnson, 1970; Breyer, 1972; Rosenbluh, 1972; Beck, 1976; Fischer, 1977; Samuel, 1977; Lennon, 1983; Barnes, 1985; Otto, 1985). This claim allows researchers to avoid explaining the often complex and contradictory findings revealed by their studies. For example, when Johnson's (1970) study on experimenter sex, verbal reinforcement, and participants' speed at a marble-dropping task fails to reveal the hypothesized experimenter sex-participant sex interaction effects, the researcher posits that this is a result of the confounding personal characteristic of experimenter attractiveness.

Sexual arousal with experimenters of the opposite sex (1970s-1980s)

Some researchers in the 1970s and 1980s propose that participants feel sexually aroused in the presence of an experimenter of the opposite sex (Harris, 1970; Johnson, 1970; O'Sullivan, 1973; Beck, 1976; Edwards, 1977; Gralton, 1979; Shilkret, 1981; Brockner, 1982; Boutcher, 1988). Some researchers posit that participants behave toward experimenters of the opposite sex in a manner similar to that which they might behave toward another person of the opposite sex around whom they feel sexual attraction. For example, a study investigating the amount of personal space needed by adolescent male participants reveals that participants express the need for more personal space with male than with female experimenters. Beck (1976) suggests that the increased heterosexual orientation of the adolescent gives the participants reason to allow female, but not male, experimenters to approach them in close proximity.

Stress and anxiety with experimenters of a particular sex (1960s, 1990s-2000s)

At distinctly different moments—in the 1960s and within past two decades—researchers propose that participants are more self-conscious, anxious, and stressed around experimenters of a particular sex, usually the opposite sex (Berkowitz, 1964; Stevenson, 1964; Ogawa, 1965; Dixit, 1969; Gold, 1969; Leventhal, 1990; Jemelka, 1991; Nilsen, 1998; Marx, 2002; Gjisbers, 2005). For example, in Jemelka's (1991) study on participants' success at solving anagrams, Jemelka conjectures that male participants do poorly with female experimenters because they feel anxious and threatened by the presence of a female authority figure.

Changed social roles of men and women (1990s-2000s)

Researchers are inclined to explain their results through the lens of time: diachronically, amid the changing social roles in the United States. Researchers hold that participants behave differently now than in earlier studies as a result of the dramatically changed roles of men and women in society (Levine 1991, Williams 1993, Kallai 2004, Reysen 2006). As women comprise more than half the college population in the United States, the continued expansion of the female sphere beyond the home has led to increasing numbers and influence of women in authority positions. Researchers suppose that participants relate to male and female experimenters in the same way that they would relate to any other male and female authority figures. Thus, as female authority figures become more commonplace and more respected, the ways participants relate to female experimenters may become increasingly similar to the ways they relate to male experimenters. For example, Reysen's (2006) study on bystander helping behavior in response to an experimenter's dropped pen hypothesized that male participants would provide more help (by handing the dropped pen to the experimenter that dropped it) than female participants, and that female experimenters would receive more help than male experimenters. Contrary to the hypothesis, there were no experimenter sex main or interaction effects and no participant sex main effects. Reysen concludes that this outcome was a result of the evolving social roles of helping: the chivalrous notion that men ought to help women is less salient than it has been historically.

Methodological limitations

Just as explanations of results have shifted over time, so too has attention to methodological constraints and problems. While some researchers point out factors that limit the reliability and/or validity of their studies, many problematic methods are left unacknowledged by the researchers. An example of a common methodological limitation that is generally *acknowledged* by researchers is the use of questionnaires relying solely on participant self-report to obtain data, limiting insofar as the participant could easily misrepresent him/herself. An example of a common methodological limitation that is generally *unacknowledged* by researchers is the over-generalization of findings: researchers frequently make claims about populations as broad as "United States residents" based on data collected with a biased sample of participants who are alike in age, background, and other characteristics. Both acknowledged and unacknowledged limiting methods are considered in this review.

The three most recurring, pressing limiting methods in the coded studies are: 1) failure to describe (and thus presumably failure to consider) experimenter demographic characteristics; 2) use of only undergraduate students as experimenters; and 3) use of only one male experimenter and one female experimenter.

Failure to consider experimenter demographics limits a study's validity in that observations of experimenter sex effects could be confounded by factors such as age, social status, race, and physical appearance. Failure to report experimenter demographics limits reliability in that future researchers will not be able to replicate the study to see if the findings hold. Researchers fail to report any demographic

characteristics (beyond sex) of the experimenters in at least one fourth of coded studies in each decade.

Undergraduate students are the only experimenters used in at least one fifth of coded studies through the 1990s. This is a serious limitation because participants are often undergraduate students as well, and a peer-peer relationship between experimenters and participants may mean that the variables do not measure what they are meant to measure, limiting a study's validity. Participants may view undergraduate students differently than they view older adults, who are likely to be seen as more credible as psychological researchers. Participants may not take an experiment seriously if it is run by a peer, and thus, the data collected by a student experimenter may be less accurate and less valid than those collected by a professor or by another older experimenter. That is, when both participants and experimenters are students, the likelihood that participants do not see experimenters as authority figures may affect studies' findings. Furthermore, if Sears (1986) is correct that undergraduate students constitute a unique subset of the population, then studies using only undergraduate students as experimenters have limited generalizability.

The use of only one male and one female experimenter is arguably the most problematic condition in experimental studies on experimenter sex effects. This condition occurs in at least one fifth of coded studies from the 1920s through the 1960s. Despite its methodological problems, the condition continues: over one third of coded studies published since 2000 use only one male and one female experimenter. With only one male and one female experimenter, there is no way to determine whether the studies' findings are a result of experimenter sex differences or

individual traits of the particular experimenters employed. Such studies are limited in both reliability and validity.

Discussion

Explanations and implications

Occurrence of studies

Researchers' willingness to pay prolonged attention to inexorable methodological quandaries is reflected by lack of support for the hypothesis that, after the second-wave feminist movement, studies on experimenter sex would decline in numbers to the levels seen prior to the movement. Experimenter sex is not a factor that can be easily extracted from empirical psychology research: even studies that avoid face-to-face experimenter-participant interactions likely retain vestiges of experimenter sex bias (for instance, in topic selection, means of data collection, and interpretations). After the plethora of studies on experimenter sex published in the 1970s suggested that experimenter sex *matters* in approximately four-fifths of studies, it would have been irresponsible for researchers to subsequently ignore the issue, despite the impossibility of entirely removing an experimenter's personal attributes and subjectivities. Thus, the steady, albeit reduced, subsequent attention to experimenter sex is encouraging. Indicative of a lasting acknowledgment of experimenter subjectivities, the continuation of studies on experimenter sex provides insight into an important aspect of the experimental paradigm that did not disappear with second-wave feminism.

Experimenter sex effects

The continued appearance of studies on experimenter sex is particularly important considering the confirmation of the second hypothesis: experimenter sex main or interaction effects are reported in over half the coded studies, including those published recently. In light of the finding that studies published since 2000 report experimenter sex effects in nearly four out of five studies, it appears that experimenter sex influence is as robust as ever. (Reports of "experimenter sex-other variable" interaction effects may be sparse because researchers often do not test for these kinds of effects.) The apparent robustness of experimenter sex effects may indicate that power relations remain very much influenced by the sexes of the parties involved despite—or perhaps in backlash against—persistent feminist efforts to break down the stereotypes and expectations largely responsible for "sexing" power plays. Alternatively, the continued observation of experimenter sex effects may reflect participants' attention (conscious or unconscious) to the personal characteristics of a given experimenter. Another possibility is that researchers and publishers have ignored Eizner's (1977), Denmark's (1988), and Unger's (2007) warnings against overemphasizing sex differences and underemphasizing similarities: research may be more likely to get published if differences rather than similarities are reported. Therefore, the studies reviewed in this meta-analysis may over-represent sex differences and under-represent similarities. It is possible that favoritism toward reported differences was not seen prior to the 1960s, explaining the finding that no coded studies published before 1960 report experimenter sex main effects. Alternatively, the finding that no coded studies from the 1920s through the 1950s

yield experimenter sex main effects may be an artifact of the small sample: only nine studies with publication dates earlier than 1960 were coded.

Experimenter sex effects with respect to specific dependent variables

The proliferation of studies on experimenter sex effects with respect to participants' task performance and/or achievement motivation has been spurred by a consistent interest in daily hierarchical interactions. The hierarchical dynamics of experimenter-participant interactions revealed by these studies have important implications for parent-child, teacher-student, mentor-mentee, and employer-employee relations. The studies' inconclusive results suggest that other aspects of the environmental context are more predictive of task performance and achievement motivation than the sex of a present authority. The findings of two other meta-analyses (Feingold 1988, Rumenik 1977), however, suggest that subdividing the research into more specific categories to test for interactions between experimenter sex and other contextual factors yields more pointed results. Feingold (1988) reports that cognitive differences between males and females (particularly with regard to a historically female deficit in mathematical and scientific cognitive abilities) have decreased over time. Rumenik (1977) deduces that children have better task performance with female experimenters, while adults (especially women) have better task performance with male experimenters. These reviews indicate that if the present analysis of studies on participants' task performance and achievement motivation had isolated and analyzed the variables 'time period' and 'participant age', interaction

effects between these variables and experimenter sex or participant sex may have been revealed.

Regarding the specifics of hierarchical relations, it has been reported that men remain more likely than women to wield social power (Powers, 2005). Therefore, it is probable that participants condition more easily and are more talkative and responsive with male than with female experimenters because—seeing a man rather than a woman in a position of authority—participants are more likely to take the experiment seriously, and to fully engage themselves. In addition, the finding that participants are more talkative and responsive with male than with female experimenters coincides with Rumenik's (1977) finding that male more than female psychotherapeutic counselors elicit information-seeking responses from patients.

Perceptions of men and women target men as more likely than women to value and expect masculinity among men (O'Neil, 1981) and femininity among women (Busk-Jensen, 1985), suggesting that participants are more likely to behave in a "sex-appropriate" manner with male than with female experimenters because participants act in accordance with what they assume is expected or desired of their behavior.

Women more often than men receive help from others (Eagly, 1986), an artifact of chivalrous notions that men ought to provide for women, and demonstrative of Glick's (1997) report that benevolent sexist attitudes persist in the United States. The finding that participants comply with and help female more than male experimenters thus remains in line with past observations and research. Women are also generally seen as more empathic than men (Hojat, 2002), and as more

oriented to interpersonal goals in group settings (Eagly, 1978). Perhaps as a consequence of women's perceived empathy and interpersonal orientation, Rumenik (1977) finds that female more than male psychotherapeutic counselors tend to elicit self-disclosure and self-expression from patients. In accordance with these findings, coded studies reveal that participants are more likely to report physical pain to female than to male experimenters: participants probably expect more empathic, interpersonal, supportive responses from women than from men.

Since self-expression tends to be more calculated in opposite-sex than in same-sex interactions (Piliavin, 1976), it seems possible that participants are more likely to be compliant, helpful, talkative and responsive, and less likely to report physical pain, with experimenters of the opposite sex than with experimenters of their own sex because participants try to impress the opposite-sex experimenter, because they are nervous, or because they are simply more engaged in the experiment than with experimenters of their own sex.

These findings indicate that children, students, and employees might be easily conditioned, might be talkative and responsive, and might behave stereotypically "feminine" if female and "masculine" if male when in the presence of fathers, male teachers, male employers, and male friends. Likewise, children, students, and employees might be particularly compliant, helpful, and self-disclosing with mothers, female teachers, female employers, and female friends. On the other hand, one might be expected to show high levels of compliance, helpfulness, talkativeness, and responsiveness when in the presence of a member of the opposite sex. Meanwhile, one might be expected to exhibit high levels of self-disclosure when in the presence

of a member of one's own sex. These contradictions might be explained by interaction effects with extraneous variables. For instance, experimenter sex might interact with participant age to predict participant talkativeness: children participants might be more talkative with male experimenters, while adolescent participants might be more talkative with opposite-sex experimenters, or vice versa.

Taken as a whole, the findings of the coded studies in this review suggest a simple answer to Sherif's (1994) query, "What can we learn from an examination of the state of [women in] psychology today that will further an equitable pursuit of knowledge?" The answer: experimenter sex effects are real and have palpable implications; thus, we can vary, study, and theorize about the causes and consequences of these effects in order to further a more equitable pursuit of knowledge.

Reported explanations of findings

Explanations of these sex-related findings have shifted over time because "[r]esearch by and about women has been influenced by the organizational and intellectual climate of the time. This context influences…how [questions] are answered and how feminist theory fits into the theoretical and disciplinary structure of psychology" (Unger, 2001, p. 3).

The climate in the United States from the 1960s through the 1970s was defined by a spirit of social and political change, one marked by three important yet inextricably bound movements: the civil rights movement, second-wave feminism, and education reform. As second-wave feminists questioned traditional assumptions

about the place of women in society, evolving perceptions of women in the sciences may have caused researchers in the 1960s to propose that participants are more self-conscious, anxious, and stressed around experimenters of a particular sex. As education reformers brought attention to the needs of the child and student, scientific researchers began to equate the role of experimental participants with that of children, pointing to the likeness between experimenter-participant interactions and parent-child interactions in reports of their studies in the 1960s and 1970s. As second-wave feminists encouraged women to broaden their spheres beyond that which was considered traditionally and stereotypically feminine, they called attention to traditional and stereotypic notions of sex-typing with respect to certain realms, triggering researchers in the 1970s to postulate that experimental participants respond to the "masculinity/femininity" of the task at hand. The civil rights movement and second-wave feminism precipitated the United States' sexual revolution of the 1970s. As societal acceptance of sexuality and sexual imagery in the media increased in the 1970s and 1980s (Levine, 2007), so did the supposition that experimental participants in psychology studies feel sexually aroused in the presence of an experimenter of the opposite sex.

The 1950s, 1960s, and 1970s also saw intensified acknowledgment of methodological contaminants in psychology research, notably with Rosenthal's (1966) critiques of experimenter bias and Guthrie's (1976) critiques of racial bias. As researchers pointed to particular contaminating factors, they explained their findings in studies on experimenter sex effects by pointing to personal characteristics of individual experimenters as a likely confounding factor.

Today, researchers' explanations of findings are largely influenced by third-wave feminism. Current third-wave feminism, with its emphasis on more idiosyncratic understandings of feminism and femininity, beckons attention to unequal relations between men and women, and between women of different backgrounds. Thus, there is a current resurgence of claims that participants might be more self-conscious, anxious, and stressed around experimenters of a particular sex. Additionally, third-wave feminists focus on the micro-politics of specific issues (such as cross-cultural notions of feminism and the morality of sex work), in consequence of the at least theoretical achievement of the more global goals of first- and second-wave feminism (such as the right to vote and the right to be free from discrimination in school and in the workplace). The immense feminist progress that characterizes the 20th century has led current researchers to explain their results through the lens of history, comparing the roles and rights of women today with those of women in the past.

Methodological limitations

The researchers who administered the coded studies varied experimenter sex to measure its effects; these researchers, therefore, are undoubtedly aware of the potentially confounding effects of experimenter demographics. In light of this fact, the repeated failure of many researchers to describe experimenter characteristics aside from sex is surprising and inexcusable. Researchers often neglect to *vary* experimenter demographics beyond sex because it generally requires extensive time and resources to do so. For instance, in order to test the effects of experimenter race

and the interaction effects between race and sex, a researcher would need to employ experimenters from every pertinent racial group (of both sexes), and to obtain comparable diversity in the participant pool. Even when logistical constraints prevent researchers from *varying* experimenter traits, however, researchers ought to at least provide demographic *descriptions* of the experimenters in their written reports in order to preclude faulty generalizations of studies' results. When researchers neglect to specify experimenter race, age, and socioeconomic status (the most salient demographic characteristics beyond sex), they contribute to the problematic assumption that the unspecified "norm" is: white, middle-aged, and upper-middle class. By failing to specify experimenter characteristics when they match the "norm", researchers perpetuate the notion that studies ought to be generalizable to a specific portion of the population, while results that pertain to other portions of the population are somehow "extra" or "peripheral", and not quite as relevant.

The use of students as experimenters is common because research is often conducted in universities where students are easily accessible and are often willing to work as volunteers or for relatively small compensation. To control for the potentially confounding effects of utilizing undergraduate experimenters, researchers need to vary the age, title, and status of experimenters, a task that would once again demand extensive time and resources. This limitation now occurs less frequently than in previous years possibly because researchers are becoming more attuned to the implicit problems; and/or researchers may now be more likely to collaborate with one another, using each other or graduate students instead of undergraduates as experimenters in studies.

The use of only one male and one female experimenter may have been a recurring limitation before the 1970s, before women began to play a significant role in psychology research, because it may have been assumed that any particular female experimenter would be interchangeable with any other. However, with the third-wave feminist emphasis on deconstructing a universal female identity and with continued acknowledgment of experimenter bias, there is no obvious justification for the continuation and recent surge in the use of only one male and one female experimenter in research on experimenter sex effects. This practice is analogous to administering an IQ test to one boy and one girl, and then, upon discovering that the girl received a higher score, claiming, “All girls are smarter than all boys.” This degree of generalizing is unjustifiable, even absurd.

Limitations and recommendations for future research

Only studies published in English are included in the meta-analysis, limiting the cross-cultural relevance of the analysis’ findings, as the results are likely only pertinent in the United States (where the vast majority of the studies were published) and perhaps only pertinent to a subset of the population in the United States. A future meta-analysis might include studies published in other languages, so as to make the sample more inclusive, and consequently, more generalizable, and to provide a basis for cross-cultural comparisons between studies on experimenter sex bias.

My personal stance as the daughter of a second-wave feminist, the subject of third-wave feminist influence, and a student at a liberal arts institution in the northeast may have affected how I chose to locate and operationally define different dependent

variables, different explanatory paradigms, and different methodological contaminants. Another researcher with a different background and stance may have drawn different distinctions between categories than I drew, and given different levels of attention to each category than I gave. A similar future meta-analysis might involve periodic consultations between several collaborating researchers to collectively make decisions regarding categorization and consideration of topics.

Study 2: Experiment

The findings of the meta-analysis led me to wonder what results replication of a past study on experimenter sex effects would yield. I chose to replicate a study conducted during the second-wave feminist surge. As the meta-analysis shows, studies published in the 1970s and studies published since 2000 report similar occurrences of experimenter sex main and interaction effects and participant sex main effects. However, researchers' explanations of findings are very different now than in the 1970s. In the 1970s, explanations of findings focused on the likeness between experimenter-participant and parent-child interactions, participants' responses to the "masculinity/femininity" of the task at hand, and participants' sexual arousal in the presence of experimenters of the opposite sex. Currently, researchers' explanations of findings focus on the changing social roles of men and women, and participants' feelings of self-consciousness, anxiety, and stress around experimenters of a particular sex. The paradoxical failure of experimenter sex effects to decline coupled with the evolution of explanations of findings (implying an increasingly progressive view of women in positions of authority) suggests that replication of an experiment conducted in the 1970s might yield similar effects to those found in the original study, but also that those effects might connote different meanings and implications than they did when the original research was published.

Method

Empirical research on experimenter sex has produced complex and mixed effects on participants' behaviors and performance. However, the results of the 11

coded studies that consider participants' ratings of experimenters (Allen 1969, Piacente 1974, Piacente 1974, Walsh 1977, Brockner 1982, Barnes 1985, Kelley 1985, Kelley 1985, Leventhal 1990, Powell 1994, Kallai 2004) along with the findings of Goldberg (1968), Paludi (1983), Swim (1989), and Kasof (1993) are consistent. They report that men are seen as more competent, accountable, professional, authoritative, and impersonal than women, while women are seen as warmer, less professional, and more likeable than men. They also generally report that experimental participants see experimenters of the opposite sex as more attractive than experimenters of their own sex. Given the apparent robustness of these findings, I decided to conduct a modified replication of one of these studies. A study conducted by Beth Stearns Piacente in 1974 was selected because it influenced the research that followed. The aim is to determine whether or not her findings remain pertinent 34 years after her research, which was conducted in the midst of second-wave feminism and well before the emergence of third-wave feminism beginning in the 1990s.

Piacente's (1974) research on participants' ratings, judgments, and perceptions of male versus female, competent versus incompetent, experimenters produced findings consistent with similar contemporaneous research (Goldberg 1968, Allen 1969, Walsh 1977, Brockner 1982, Paludi 1983, Barnes 1985, Kelley 1985, Kelley 1985, Swim 1989, Leventhal 1990, Kasof 1993, Powell 1994, Kallai 2004). Piacente reports that: competent male and female experimenters are rated equally, while incompetent female experimenters are judged more harshly than their incompetent male counterparts, and competent female experimenters are seen as less feminine than incompetent female experimenters. She postulates that participants

expect incompetence of women, and equate competence with masculinity and incompetence with femininity. This specific focus on participants' judgments of male versus female, competent versus incompetent, experimenters has not been examined since the 1970s and therefore, it is worth determining whether or not participants' perceptions and judgments have changed over the past 34 years.

I hypothesized that my modified replication of Piacente's study would produce similar results to those found in 1974. Specifically, the hypotheses stated that: 1) competent male and female experimenters would be rated equally, while incompetent female experimenters would be judged more harshly than incompetent male experimenters; and 2) competent female experimenters would be seen as less feminine than incompetent female experimenters. There are three grounds for these hypotheses. First, while explanations of findings in studies on experimenter sex effects have changed over time, the findings themselves have not demonstrated a clear evolution. Second, personal observations and experiences indicate that undergraduate students maintain a tendency to rate female authority figures more harshly than male authority figures. Finally, while women have become more welcome in spheres of authority and influence with a decrease in hostile sexism noted by Glick (1997), the notions and implications of traditional femininity seem to largely remain with Glick's noted continuance of "benevolent" sexism.

The results of this empirical replication are important: as the first study to examine this issue in several decades, the findings will indicate whether we need to reassess the ways in which people rate and judge males and females performing at different levels of competence. Whether the hypotheses are confirmed or

disconfirmed, the findings can have important implications for “real world” male-female interactions. If the hypotheses are confirmed and Piacente’s findings are reproduced, then we can suppose that before reaching a level of unquestioned competence, females must still struggle more than similarly capable males to gain recognition and respect in a given field. Furthermore, we can assume that, by achieving a level of unquestioned competence, females may be forced to sacrifice (for better or for worse) their femininity in the eyes of others. If, on the other hand, the hypotheses are disconfirmed and Piacente’s findings are not reproduced, we can consider that views toward women have, in fact, evolved. We can assume that, even before reaching a level of unquestioned competence, women are now judged relatively equally with their male counterparts. We also can assume that a woman need not sacrifice her femininity (from outsiders’ perspectives) with the achievement of competence in a chosen field.

In Piacente’s (1974) study, undergraduate introductory psychology students (University of South Florida) were divided into 16 groups, each of which viewed one of eight videos portraying an experimenter (a graduate or undergraduate student) conducting an experiment in which participants (undergraduate students) delivered or prevented the delivery of electric shocks to a rat. The experimenter in each film was one of two men or two women, each of whom acted competently in one film and incompetently in the other. The administrators of the viewing sessions were two men and two women, each of whom administered four sessions. Participants were asked to rate the experimenter they saw in the film on a series of semantic differential scales (Piacente, 1974).

Due to current ethical standards, it was not appropriate to replicate Piacente's study with the same premise of delivering real or feigned electric shocks to rats. Instead of an experiment on a rat's physiological responses to electric shocks, I filmed an experiment on a person's physiological responses (as measured by an electrocardiogram) to apprehension regarding the receipt, denial, or removal of a monetary sum. The four experimenters in the films—each of whom acted competently in one film and incompetently in another—were four white undergraduate students, two females and two males, all of whom were recruited from other colleges. White undergraduate students were employed as experimenters to match the demographic characteristics of the experimenters employed by Piacente. As a result of limited time and resources preventing the utilization of a wider array of demographic characteristics among experimenters, this study sacrifices some degree of validity for the sake of reliability.

The four experimenters acted as participants in the films in which they were not experimenters (since each true participant would only view one film), along with two other female undergraduate students, one from Bentley College and the other from Wesleyan University. The scripts were matched as closely as possible (considering the altered premise of the experiment) with those used by Piacente (Penner, 1973). The scripts followed by the experimenters and participants in the films, for both the competent and incompetent conditions, can be found in Appendix B. The experimenters and participants in the films were blind to the purpose and hypotheses of the study. The films were made in a standard classroom at Bentley College. Each film was 11-16 minutes in duration. The films were transferred from a

cassette onto a computer using I-Movie software, and then from a computer onto two DVDs using I-DVD software.

The study was approved by the Wesleyan University Psychology Ethics Committee. Participants were students in two introductory psychology courses (PSYC 105 first semester and second semester). Four other white undergraduate students (three males and one female) were recruited to help me administer the viewing sessions. Once again, white undergraduate students (including myself) were employed as administrators in order to match the demographic characteristics of the administrators employed by Piacente. Participants randomly assigned to my research signed up for one of eight time slots first semester and one of ten time slots second semester to participate. Each group comprised 5-14 participants. There were eight cells for each of eight films presented: each cell comprised 17-26 participants for a total of 166 participants. Each participant viewed only one film and is thus included in only one cell. With the exception of myself, all administrators were blind to the study's hypotheses.

The sessions were held in one of two small classrooms equipped with chairs, a DVD player, and a monitor. When the participants entered the room, the administrator handed each of them two copies of a consent form: one for them to read, sign and return; and one for them to keep (see Appendix C). The administrator then read the following instructions (matched nearly verbatim with those read by the administrators in Piacente's study):

> We are in the process of training some people to run an experiment for us. We have filmed them in the process of running the experiment, and we need some opinions as to their

> competence. The experiment you will be viewing is designed to study heart rate and apprehension. Some of the participants were confederates as our only interest at this time lies in the experimenter's performance. The experimenter was not acquainted with the participants. You will notice a few cuts in the film. We felt that the setting up of the experiment, the practice trials, and the last few trials would give you a fair representation of the performance of the experimenter. Don't be alarmed by the incident on trial 17, the participant's heart rate almost immediately returned to normal. I would like for you to view this film now and then fill out a questionnaire which will tell us what kind of job you think the experimenter is doing. The experimenter will receive your feedback (which will remain anonymous) and then repeat the experiment with different participants. We are interested in how the experimenter's performance will change after receiving your feedback. Remember, we are interested in your judgment of the experimenter's behavior. Please pay attention to that.

After the Psychology 105 participants viewed the film, the administrator stopped the DVD and read the following:

> We want you to rate the experimenter in the film you just saw on each of the scales on the following pages. In filling out this questionnaire, please make your judgments on the basis of what these things mean *to you.* On each page of this packet you will find a different concept to be judged and beneath it a set of scales. You are to rate the concept on each of these scales in order. The direction toward which you check, of course, depends upon which of the two ends of the scale seems most characteristic of the experimenter. If you think both sides of the scale are equally

associated with the experimenter or if you think the scale is completely irrelevant, then you should place your check-mark in the middle space.

Three important notes for you to remember are:

1) Place your check-marks *in the middle of spaces*, not on the boundaries, crossing between spaces.
2) Be sure to check every scale for every concept—*do not omit any*.
3) Never put more than one check-mark on a single scale.

Make each item a separate and independent judgment. Work at a fairly high speed through this test. Do not worry or puzzle over individual items. It is your first impressions, the immediate "feelings" about the items that we want. On the other hand, please do not be careless, because we want your true impressions.

The administrator then distributed a questionnaire asking participants to provide demographic information about themselves and to rate the experimenter in the film they just viewed on 20 semantic differential scales, which extracted participants' perceptions of the filmed experimenter's performance, potency, activity, and competency (see Appendix D). The semantic differential scales were the same as those used by Piacente. When all the participants had completed the questionnaire, the administrator collected them and then distributed a debriefing form explaining the true nature of the research (see Appendix E). The administrator concluded the session, asked the participants not to mention anything about the study to anyone else, and thanked them for their participation.

Data analysis

The results of the participants' questionnaire ratings were entered into Microsoft Excel and then transferred into SPSS. Analyses of descriptive statistics of each semantic differential scale and a series of independent samples t-tests to find statistically significant results were run. The independent samples t-tests determined the main effects of experimenter condition (competent versus incompetent), experimenter sex, administrator sex, participant sex, individual experimenter, and individual administrator. Splitting the file by each variable and then running independent samples t-tests for each remaining variable yielded the interaction effects of experimenter condition-experimenter sex, experimenter condition-administrator sex, experimenter condition-participant sex, experimenter sex-administrator sex, experimenter sex-participant sex, and administrator sex-participant sex. Every main and interaction effect revealed by statistical analyses is reported. Isolating the file by conditions of each variable, and then splitting the file by another variable, and running independent samples t-tests for each remaining variable in turn yielded second-order interaction effects of experimenter condition-experimenter sex-administrator sex, experimenter condition-experimenter sex-participant sex, experimenter condition-administrator sex-participant sex, and experimenter sex-administrator sex-participant sex. Since analyses of second-order interaction effects revealed such a multitude of findings, only those of particular import that were not revealed by analyses of main effects or first-order interaction effects are reported.

The results are compared with those found by Piacente. The most salient (robust and important) main, interaction, and second-order interaction effects are

extracted from the results and put forth for further discussion. In Piacente's data analysis, participants were randomly eliminated until there were equal numbers in each cell, for a total of 214 participants, and the level of significance used was p=.025. The present study does not eliminate participants, but instead allows for different numbers of participants per cell, and the level of significance used is p=.05. The present study differs from Piacente's methods in these two domains because of a smaller sample. If participants had been eliminated from every cell except the smallest (containing 17 participants), 136 rather than 166 participants would have been included in the data analyses, limiting the study's reliability. In addition, some effects revealed in Piacente's study may not be similarly revealed in a study using fewer participants and the same level of significance (p=.025). Therefore, because this study uses about 50 fewer participants than Piacente used, it is necessary to use a higher level of significance (p=.05).

Results

Main effects

Experimenter condition main effects

Analyses of independent samples t-tests on experimenter condition (competent versus incompetent) reveal that participants rate experimenters (in the films) in the competent condition as *better*, more *successful*, more *positive*, *wiser*, *harder*, *stronger*, more *severe*, more *serious*, *calmer*, more *deliberate*, more *confident*, more *professional*, more *efficient*, more *experienced with the instruments*, and more *competent* than experimenters in the incompetent condition. The main effects of experimenter condition are similar to Piacente's findings in that participants rate

competent experimenters differently from incompetent experimenters in about 75% of the 20 semantic differential items. Thus, participants reliably distinguish between competent and incompetent experimenters.

Experimenter sex main effects

Piacente found that male experimenters (regardless of competency level) were seen as *better*, less *excitable*, less *attractive*, and more *masculine* than female experimenters. The two latter findings—that male experimenters were seen as less *attractive* and more *masculine* than female experimenters—are replicated in the present research. However, male experimenters are not seen as *better* or as less *excitable* than female experimenters; however, they are seen as *harder* and more *nervous*.

Administrator sex main effects

Piacente did not comment on the main effects of administrator sex. In the present study, participants rate experimenters as more *efficient* with male than with female administrators.

Participant sex main effects

Piacente found that male participants rated experimenters as more *severe* than did females. The present research, on the other hand, finds that female participants rate experimenters as *better*, *wiser*, more *serious*, more *complex*, more *professional*, more *efficient*, more *experienced with the instruments*, and more *competent* than do

male participants. Evidently, female participants rate experimenters more positively than male participants.

Individual Experimenter Effects

In Piacente's study, individual taped experimenters differed only on the bipolar adjective *successful-unsuccessful*. In the present research, in contrast, individual experimenters differ from one another on five bipolar adjectives: *hard-soft*, *confident-nervous*, *masculine-feminine*, *strong-weak*, and *serious-humorous*.

Upon analyses of individual experimenter differences, it becomes clear that several of the experimenter sex main effects are at least partly attributable to a characteristic of a specific experimenter. The finding that male experimenters are seen as *harder* than female experimenters is partly attributable to the fact that there is a large discrepancy between ratings of one particular male experimenter and one particular female experimenter. The finding that female experimenters are seen as more *confident* than male experimenters is largely attributable to the fact that one male experimenter is perceived as more *nervous* than any of the other three experimenters. The finding that male experimenters are seen as more *masculine* than female experimenters is partly attributable to the fact that one female experimenter is seen as more *feminine* than any of the other three experimenters. Other differences between individual experimenters include; one male experimenter is seen as *stronger* and more *humorous* than the other male experimenter, and one male experimenter is seen as *stronger* than one female experimenter.

Individual administrator effects

To discern the potential confounding effects of individual administrators, independent samples t-tests comparing administrators reveal differences on 11 bipolar adjectives. Participants rate experimenters as *harder* with administrator #2 (male) than with administrator #3 (female) or with administrator #4 (male). The finding that male administrators elicit higher ratings of *efficiency* than female administrators is partly attributable to a large discrepancy between the ratings of *efficiency* elicited by one male administrator (#5) versus those elicited by one female administrator (#3). One administrator (#4) elicits higher ratings of *simplicity* than administrators #1, #3 and #5. One female administrator (#1) elicits lower ratings of *excitability* than the other female administrator (#3), while one male administrator (#5) elicits higher ratings of *masculinity*, *humor*, *excitability*, and *simplicity* than another male administrator (#4). There are notable differences between participants' ratings of experimenters with administrator #3 (female) and administrator #4 (male): participants rate experimenters as more *humorous*, more *excitable*, less *professional*, and less *experienced with the instruments* with administrator #3 than with administrator #4. Participants rate experimenters as more *active* and more *efficient* with administrator #5 (male) than with administrator #1 (female). Participants see experimenters as more *attractive*, *harder*, and *stronger* with administrator #5 (male) than with administrator #2 (male).

Statistical significance of main effects

Table 1 shows the significance (the ‘p’ value) of every statistically significant result (p < or = .05) revealed by analyses of main effects. ‘P’ values are rounded to the nearest thousandth. When numerous ‘p’ values appear in one cell, they are listed in the order in which they are reported.

Table 1

	E condition	E sex	A sex	P sex	Individual E	Individual A
good-bad	.000			.006		
successful-unsuccessful	.000					
positive-negative	.000					
wise-foolish	.000			.034		
attractive-unattractive		.000				.041
hard-soft	.001	.012			.004	.022, .011, .010
masculine-feminine		.000			.000, .005, .000	.018
strong-weak	.000				.019, .014	.031
severe-lenient	.000					
serious-humorous	.000			.051	.003	
fast-slow						.030, .006
active-passive						.032
excitable-calm	.000					.023, .027, .015
impulsive-deliberate	.000					
simple-complex				.027		.014, .028, .001, .001
confident-nervous	.000	.016			.009, .003, .001	

professional-novice	.000			.002		.053
efficient-inefficient	.000		.015	.045		.033, .034
experienced-inexperienced with the instruments	.000			.016		.033
competent-incompetent	.000			.001		
excitable-calm	.000					.023, .027, .015
impulsive-deliberate	.000					
simple-complex				.027		.014, .028, .001, .001
confident-nervous	.000	.016			.009, .003, .001	
professional-novice	.000			.002		.053
efficient-inefficient	.000		.015	.045		.033, .034
experienced-inexperienced with the instruments	.000			.016		.033
competent-incompetent	.000			.001		

Interaction effects

Experimenter condition-experimenter sex interaction effects

Piacente found that male incompetent experimenters were seen as *calmer*, *harder*, and *stronger* than female incompetent experimenters, while there were no such differences between ratings of male and female experimenters in the competent condition. The present research reproduces only one of these findings: incompetent male experimenters are seen as *calmer* than incompetent female experimenters.

Incompetent male experimenters are also seen as more *serious* and more *passive* than incompetent female experimenters. Meanwhile, competent male experimenters are seen as *harder*, *simpler*, and more *nervous* than competent female experimenters. These findings disconfirm the first hypothesis that competent male and female experimenters would be rated equally, while incompetent female experimenters would be judged more harshly than incompetent male experimenters.

Piacente also found that competent female experimenters were seen as more *severe* than any other experimenters, and as *stronger*, *harder*, and less *feminine* than incompetent female experimenters, while there were no comparable differences between competent and incompetent male experimenters. The present research, on the other hand, finds only one difference specifically between female experimenters in the competent condition and female experimenters in the incompetent condition: competent female experimenters are seen as more *complex*. Competent male experimenters are seen as *harder* and more *active* than incompetent male experimenters. These findings disconfirm the second hypothesis that competent female experimenters would be seen as less feminine than incompetent female experimenters.

Experimenter condition-administrator sex interaction effects

Piacente found that in viewing sessions run by male administrators, participants saw experimenters in the competent condition as more *masculine* than experimenters in the incompetent condition. The present research, on the other hand, finds that only in the competent experimenter condition, participants see

experimenters as more *successful*, more *attractive*, more *efficient*, and more *competent* with a male administrator than with a female administrator. Meanwhile, only in the incompetent experimenter condition, participants see experimenters as more *successful*, more *attractive*, and more *humorous* with a female administrator than with a male administrator. Only with female administrators, participants see experimenters as more *attractive* in the incompetent condition than in the competent condition. Thus, participants perceive competent experimenters more positively with male than with female administrators, while participants perceive incompetent experimenters more positively with female than with male administrators.

Experimenter condition-participant sex interaction effects

Further analyses of experimenter condition effects show that only female participants rate competent experimenters as *better*, more *complex*, more *confident*, more *professional*, and more *competent* than incompetent experimenters. Only female participants see incompetent experimenters as more *severe* than competent experimenters. Thus, female participants' ratings are more affected by experimenter condition than male participants' ratings.

Experimenter sex-administrator sex interaction effects

Only with male administrators, participants see female experimenters as more *efficient* than male experimenters, and participants see male experimenters as *harder*, *slower*, and more *nervous* than female experimenters.

Experimenter sex-participant sex interaction effects

Piacente found that only male participants rated male experimenters as more *efficient* than female experimenters. The present research reveals different experimenter sex-participant sex interaction effects: only female participants rate male experimenters as *better*, *wiser*, *harder*, more *masculine*, and more *competent* than female experimenters. Female participants also rate female experimenters as more *serious*, *calmer*, and more *professional* than male experimenters. Only male participants see female experimenters as *faster* than male experimenters.

Administrator sex-participant sex interaction effects

Further analyses of interaction effects show that only with male administrators, female participants see experimenters as more *professional* than do male participants. Only with female administrators, female participants see experimenters as *better*, *calmer*, more *complex*, more *efficient*, and more *experienced with the instruments* than do male participants.

Statistical significance of interaction effects

Table 2 shows the significance (the ‘p’ value) of every statistically significant result ($p <$ or $= .05$) revealed by analyses of interaction effects. ‘P’ values are rounded to the nearest thousandth. When numerous ‘p’ values appear in one cell, they are listed in the order in which they are reported.

Table 2

	E condition-E sex	E condition-A sex	E condition-P sex	E sex-A sex	E sex-P sex	A sex-P sex
good-bad			.006		.018	.010
successful-unsuccessful		.021, .047				
positive-negative						
wise-foolish					.008	
attractive-unattractive		.052, .023, .012				
hard-soft	.039, .004			.005	.010	
masculine-feminine					.035	
strong-weak						
severe-lenient			.000			
serious-humorous	.044	.001			.031	
fast-slow				.033	.022	
active-passive	.004, .002					
excitable-calm	.002				.030	.023
impulsive-deliberate						
simple-complex	.032, .028		.022			.002
confident-nervous	.017		.025	.005		
professional-novice			.019		.006	.010
efficient-inefficient		.005		.011		.025
experienced-inexperienced with the instruments						.041
competent-incompetent		.033	.003		.003	

Second-order interaction effects

Experimenter condition-experimenter sex-administrator sex interaction effects

Piacente found that when female administrators presented films of female experimenters, participants did not discriminate between competent and incompetent conditions on the bipolar adjective *impulsive-deliberate*. In the present study, male and female administrators elicit different ratings of *female* experimenters in the competent condition (participants give higher ratings of *complexity*, *efficiency*, and *softness* with male administrators than with female administrators), but participants' ratings of *male* experimenters in the competent condition are the same with male versus female administrators. Again, it appears as though ratings of male experimenters are more consistent across conditions than ratings of female experimenters, which are affected by administrator sex. Participants give higher ratings of *femininity* to incompetent female experimenters than to competent female experimenters, only with female administrators. Therefore, repudiation of the hypothesis that incompetent female experimenters would be seen as more feminine than competent female experimenters must be at least partially renounced, as the hypothesis is apparently supported when the administrator is female.

Experimenter condition-experimenter sex-participant sex interaction effects

Piacente found that male (but not female) participants rated incompetent female experimenters as *softer* than competent female experimenters. In the present study, while analyses of participant sex main effects show that female participants rate experimenters more positively than male participants, analyses of second-order

interaction effects reveal that this is not the case regarding incompetent female experimenters, toward whom there are no differences between male and female participants' ratings. Female (but not male) participants give higher ratings of *foolishness* to incompetent female experimenters than to incompetent male experimenters. Thus, interestingly, the hypothesis that participants would rate incompetent female experimenters more harshly than incompetent male experimenters is supported only by female participants with respect to one bipolar adjective: *wise-foolish*. Female participants give higher ratings of *calmness* and *confidence* to competent female experimenters than to competent male experimenters, while giving higher ratings of *foolishness*, *excitability*, and *softness* to incompetent female experimenters than to incompetent male experimenters. These findings intimate that female participants' ratings of female experimenters are more based on experimenter performance than female participants' ratings of male experimenters, which remain more stable across competency levels.

Experimenter condition-administrator sex-participant sex interaction effects

Only with male administrators do female participants give higher ratings of *attractiveness* to competent experimenters than to incompetent experimenters. Female (but not male) participants appear to associate attractiveness with competency, with male (but not with female) administrators. The tendency for female participants to give more positive ratings than male participants does not maintain with regard to incompetent experimenters presented by male administrators.

Experimenter sex-administrator sex-participant sex interaction effects

The tendency for female participants to give more positive ratings than male participants does not maintain with regard to male experimenters presented by female administrators or with regard to female experimenters presented by male administrators. Female participants rate experimenters more positively than male participants only when experimenter sex and administrator sex are the same. Male (but not female) participants give higher ratings of *complexity* and *efficiency* to female experimenters with male administrators than with female administrators. Again, perceptions of male experimenters seem to remain comparatively consistent across conditions, while perceptions of female experimenters are more variable: in this case, affected by administrator sex.

Statistical significance of second-order interaction effects

Table 3 shows the significance (the 'p' value) of every statistically significant result (p < or = .05) included in the description of second-order interaction effects. 'P' values are rounded to the nearest thousandth. When numerous 'p' values appear in one cell, they are listed in the order in which they are reported.

Table 3

	E condition-E sex-A sex	E condition-E sex-P sex	E condition-A sex-P sex	E sex-A sex-P sex
good-bad				
successful-unsuccessful				
positive-negative				
wise-foolish		.003		

attractive-unattractive			.010	
hard-soft	.036	.021		
masculine-feminine	.031			
strong-weak				
severe-lenient				
serious-humorous				
fast-slow				
active-passive				
excitable-calm		.005, .041		
impulsive-deliberate				
simple-complex	.031			.002
confident-nervous		.047		
professional-novice				
efficient-inefficient	.013			.007
experienced-inexperienced with the instruments				
competent-incompetent				

Discussion

The most salient (robust and important) findings will be discussed. These findings include the main effects that: 1) male experimenters are not seen as *better* than female experimenters; 2) female participants rate experimenters more positively than male participants; and 3) participants give notably disparate ratings of individual

experimenters and with individual administrators. Interaction effects to be discussed include: 1) ratings of competent experimenters are more positive with male than with female administrators; 2) ratings of incompetent experimenters are more positive with female than with male administrators; 3) ratings indicate that incompetent experimenters are seen as more attractive than competent experimenters with female administrators; and 4) ratings by female participants mark male experimenters as better, wiser, and more competent than female experimenters. Second-order interaction effects to be discussed include: 1) female participants see incompetent female experimenters as more foolish than incompetent male experimenters; partly confirming the hypothesis that competent male and female experimenters would be rated equally, while incompetent female experimenters would be judged more harshly than incompetent male experimenters; 2) female participants' ratings along with perceptions of female experimenters vary more greatly with context than male participants' ratings along with perceptions of male experimenters, which remain comparatively stable; and 3) female participants appear to associate attractiveness with competence, with male (but not with female) administrators, while male participants do not.

Explanations and implications

Analyses of experimenter sex main effects reveal that male experimenters are not seen as *better* than female experimenters, contradicting Piacente's findings as well as those of the several other studies which support the Golberg paradigm that identical works or performances presented by a "man" are rated more positively than

when presented by a “woman” (Walsh 1977, Paludi 1983, Swim 1989, Kasof 1993). However, analyses of second-order interaction effects reveal that female participants do, in fact, confirm the Goldberg paradigm. Female participants see male experimenters as better, wiser, and more competent than female experimenters. This may be because females think that they are expected to rate women more highly than men, and thus, they overcompensate by doing the opposite. Alternatively, females may be socialized to view other women in a competitive light, increasing their denigration of them; and to view men in a way that reflects potential attraction, judging them favorably. Another possibility is that female participants have internalized sexist notions that they perceive around them, inducing them to truly view male experimenters as better, wiser, and more competent than female experimenters. The fact that male participants make no such distinctions indicates that, if females have internalized sexist notions that they perceive around them, such perceptions may not be justified. Male participants’ disconfirmation of the Goldberg paradigm suggests a promising decrease of negative stereotypes (among men) about women.

The tendency for female participants to rate experimenters more positively than male participants coincides with the common conjecture that females are either naturally or socialized to be generous and forgiving, while males are either naturally or socialized to be competitive (Wojciszke, 2003). This finding suggests that in hierarchical organizations such as school, the workplace, or the family, subordinate females may be more likely than subordinate males to assess super-ordinate figures positively and to treat them accordingly. Taking heed of Eizner’s (1977), Denmark’s

(1988) and Unger's (2007) precautions against magnifying sex differences while ignoring sex similarities, it must be acknowledged that while female participants rate experimenters more positively than male participants on eight bipolar adjectives, there are no differences between female and male participants' ratings of experimenters on the other twelve scales. More than half the time, males and females agree.

The common association between femininity and generosity and forgiveness coupled with the common association between masculinity and competitiveness also illuminates the finding that participants perceive competent experimenters more positively with male administrators and incompetent experimenters more positively with female administrators. The presence of a male administrator seems to have triggered participants' appreciation of and respect for experimenter competence, while the presence of a female administrator appears to have enforced participants' forgiveness of experimenter incompetence. One aspect of this finding that seems particularly notable is that participants see experimenters as more attractive in the incompetent condition than in the competent condition with female (but not with male) administrators, implying that femininity suggests an association between incompetence and attractiveness.

We must be careful in assuming the implications of these findings, as it is *not* the case that participants rate competent male experimenters more positively than competent female experimenters or incompetent female experimenters more positively than incompetent male experimenters. It is also *not* the case that incompetent female experimenters are seen as more attractive than competent female

experimenters. Thus, it is unlikely that people perceive competence in males and incompetence in females positively or that people see incompetent females as attractive. Rather, people seem more likely to *acknowledg*e competence in the *presence* of males and to *forgive* incompetence in the *presence* of females, and to acknowledge attractiveness even in the face of incompetence when in the midst of a forgiving female presence. These findings offer, for instance, one answer to the debate over whether Bill Clinton's ambiguously effective assistance is more of a help or hindrance to Hillary Clinton's public image: if male presence does, in fact, enhance perceptions of competence, she might benefit from his support.

A related second-order interaction effect is that female (but not male) participants appear to associate attractiveness with competence, with male (but not with female) administrators. Once again, male administrators elicit more positive ratings than female administrators in the competent experimenter condition. The particular finding about attractiveness ratings, however, is important in that it only pertains to female participants, implying that male administrators intimate a connection between competence and attractiveness only for females. This finding implies that females associate competence with attractiveness in men. Once again, however, we must be careful in our assumptions regarding this finding. As female participants did *not* rate male experimenters in the competent condition as more attractive than male experimenters in the incompetent condition, it does not appear as though females are necessarily attracted to competence in men, but simply that they *associate* competent men with attractiveness.

There are notable differences between ratings of individual experimenters and ratings elicited by individual administrators, indicating that characteristics of individuals (as suggested by researchers in the 1980s) do, in fact, influence participants more significantly than the isolated factor sex: participants are more attuned to experimenters' and administrators' overall self-presentations than to the overarching characteristic sex. Individual administrators elicit far more variability in ratings than individual experimenters, further suggesting that increased immediacy (via salience through proximity or some other means) to the participants increases the influence of personal traits. Piacente, on the other hand, found only one individual experimenter effect, suggesting that participants in 1974 were more affected by experimenter sex than by experimenters' specific self-presentations. This discrepancy between Piacente's findings and the present study's findings invites the encouraging thought that people are becoming less and less likely to judge others based on sex and more and more likely to judge others as a whole. The variability in the ratings of individual experimenters and administrators raises concerns about studies on experimenter sex effects that use only one male and one female experimenter.

While female participants tend to be relatively generous in their ratings, they are less forgiving of *foolishness* in incompetent females than in incompetent males. This vilification may arise out of worry or resentment that women behaving foolishly and incompetently might reflect poorly on females in general. This potential worry is understandable in light of the fact that the women in our society who receive the most publicity and attention include Paris Hilton, Lindsay Lohan, and Britney Spears, whose follies and mishaps people follow and condemn with apparently avid interest.

By denigrating incompetent female experimenters' *foolishness*, female participants may be attempting to protect against acceptance and expectation of foolishness in girls and women at large. As more and more women like Hillary Clinton (who has lately been receiving more public attention than any socialite or pop star) come into the limelight, females' perceptions of what others expect from women might evolve, and female forgiveness of incompetence might extend even to "foolish" women. It is also possible, as mentioned previously, that female participants have internalized sexism, resulting in harsher criticisms of foolishness in women than in men. Male participants, on the other hand, do *not* rate incompetent female experimenters as more *foolish* than incompetent male experimenters, encouragingly insinuating that: 1) potential female fears regarding all-encompassing perceptions of foolish incompetent women are not necessarily justified; and 2) females potentially perceive and internalize more sexism than males express. The finding of no differences in male participants' judgments of incompetent male versus female experimenters departs from Piacente's findings and defies the hypothesis, implying that presently—more than in 1974 when Piacente conducted her research—men and women are held to similar standards of performance.

The finding that competent female experimenters are seen as less feminine than incompetent female experimenters in the presence of female administrators indicates a disappointing perseverance of the association between competence and masculinity, and incompetence and femininity. This association is tempered by the presence of male administrators, possibly because they provide a masculine basis for comparison: in contrast with male administrators, competent and incompetent female

experimenters alike are seen as markedly feminine. These findings suggest that if several female job applicants were assessed by potential employers in a group interview, the most competent applicants would be seen as more masculine than the others. However, if the group of applicants included males as well as females, the competent females would be seen as feminine. Thus, when competing with other females (but not when competing with males), women might be forced to sacrifice some degree of perceived femininity in order appear competent. These findings coincide with Heilman's (2001) observation that "[w]omen who prove to be competent and to have succeeded at 'male' work violate [a] normative prescription and therefore arouse disapproval and are penalized; they are regarded very differently than men who engage in precisely the same behavior" (p. 671). The association between competence and masculinity, and incompetence and femininity, is not as strong as found in Piacente's study (in which the association was revealed with male as well as female administrators), indicating some deterioration of this link. However, its partial endurance is concerning.

Also indicative of unequal perceptions of sex is the finding that ratings of female experimenters are more variable depending upon context than ratings of male experimenters, which remain comparatively consistent across competency and administrator sex conditions. These findings intimate that participants hold stable preconceived notions about what to expect from male experimenters while possessing no such expectations of female experimenters. Assumptions about male authority figures might be difficult to alter, while opinions on female authority figures may be more mutable. It might be easier for female than male authority figures to gain and

lose esteem based on factors beyond sex. This discrepancy might be a result of the fact that people are more accustomed to seeing men in positions of leadership than women (especially in relation to scientific endeavors). As a result, people might think they know what to expect from men, but not from women, in such positions.

The above two findings—that the association between masculinity and competence, and femininity and incompetence, at least somewhat endures, and that perceptions of female experimenters are more variable than perceptions of male experimenters—reflect the fact that while women continue to make social gains, the uppermost echelon continues to be governed by men. While a greater number of women than men now attend college, men continue to receive higher salaries than women and to hold positions as CEOs (Blau, 2000), and while women are equipped with the skills and opportunities to run for elective office, governmental bodies are still dominated by men (Fox, 2001). More changes are required to untangle these associations between sex and competence. The above two findings are of particular interest in light of the current primary race for the 2008 democratic presidential candidate nomination. Media and public responses to Hillary Clinton have reinforced the partial endurance of the association between masculinity and competence, and femininity and incompetence: Clinton—rarely accused of incompetence—has often been the target of remarks (sometimes negative and sometimes neutral, but never positive, in tone) for her allegedly “masculine” image. More subtly, the finding that perceptions of female experimenters are more variable than perceptions of male experimenters has been repeatedly reflected through Clinton’s campaign against Obama. For example, Clinton has suffered more for denigrating her opponent than

has Obama. She has been declared aggressive, vicious, even a “monster” while like attacks from the Obama camp have received minimal attention in comparison. More pointedly, public perception of Clinton appears to be more greatly affected by extraneous variables than public perception of Obama: the media has repeatedly deemed Clinton’s attire and physical appearance worthy of report, as if there is any valuable knowledge to gain from such information, while the superficial aspects of Obama’s self-presentation have gone comparatively unnoticed.

The patterns around perceptions of female and male experimenters are mirrored by the patterns around judgments by female and male participants. Ratings by female participants are more variable than ratings by male participants, which remain comparatively consistent across competency and administrator sex conditions, hinting that females are more in tune with and affected by contextual cues than males, who seem more likely to hold preconceived notions about what to expect from an experimenter or administrator. Once again, females more than males may fear judgment based on immediately salient traits. In turn, they may be less likely than males to jump to conclusions about those they are judging. Alternatively, it is plausible that female participants were more engaged in the experiment and spent more time thinking about their questionnaire responses than male participants as a result of the previously mentioned implication that females are likely to assess and respect authority figures (in this case, the administrators) more highly than males. Female participants may have taken the study more seriously than male participants, and produced more complex data.

The findings of both Study 1's meta-analysis and Study 2's empirical replication reiterate the conclusions of Gannon's (1992) literature review: "Although the present data are encouraging in that they do indicate change, they are rather discouraging in that we still have a moderately sexist discipline" (p. 395). Gannon's observation of a "gradual ideological shift toward equality between women and men" (p. 394) from 1970 to 1990 has apparently continued. However, there is room for more progress.

Limitations and recommendations for future research

In spite of the criticism of using only one male and one female experimenter, using two male and two female experimenters accompanied by two male and two female administrators is, admittedly, not a huge correction. Although the present study's design matched that of Piacente, a greater number of experimenters and administrators would increase the study's validity.

A relatively homogenous sample also limits the reliability of the present study's findings. All participants are introductory psychology students at Wesleyan University. A future study might use a more representative sample, increasing the study's reliability and generalizability. Similarity in demographic characteristics of experimenters (white undergraduate students from the Boston area, including three business students), administrators (white undergraduate psychology students from Wesleyan) and participants (introductory psychology students from Wesleyan) constrains the generalizability of the present study's findings. Future studies should incorporate a wider array of experimenters, administrators, and participants,

particularly with regard to the traits most notably shared by those in the present study: age, race, and educational background. Augmenting demographic diversity in each group would improve the study's reliability and external validity.

Although it is possible, as some of the findings intimate, that individual characteristics of experimenters and administrators are truly more influential than sex, it is also possible that experimenters and administrators failed to sufficiently standardize their delivery, and that each person comported himself/herself in a particularly distinct manner, potentially obscuring the data. Future studies should rehearse experimenter and administrator performance. By achieving nearly identical performances from each experimenter and each administrator, the researcher could more effectively manipulate and measure the effects of experimenter and administrator sex. (Another way to control for individual differences in delivery would be to use a greater number of experimenters and administrators.)

The study's reliability may have suffered as a consequence of ethical concerns preventing a more exact replication of Piacente's research. Had the filmed experiments been on a rat's response to electric shock instead of on a person's response to a situation facilitating apprehension, participants' perceptions of the experimenters may have differed. Specifically, it is possible that participants would judge experimenters of one sex but not the other more negatively for inflicting moderate physical pain upon a helpless animal than for inflicting mild emotional distress upon a person.

A final recommendation for future research is to conduct related field studies to determine whether or not this study's findings pertain to situations beyond the

laboratory. For example, it would be interesting to replicate Basow's (1987) study on college students' perceptions of male versus female professors, which found that male professors were rated more positively than female professors. A related field study today would ascertain the extent to which Basow's results remain prevalent over 20 years later, and the extent to which the present study's findings are applicable beyond the laboratory.

Appendices

Appendix A

Key: E=experimenter, S=subject, P=participant, M=male, F=female, SOE=sex of experimenter, SOP=sex of participant, psyc=psychology, NI=not included

Author: Wayne H. Holtzmann
Title: The Examiner as a Variable in the Draw-A-Person Test
Date: 1952
Journal: Journal of Consulting Psychology
Volume: 16 (2)
Pages: 145-148

Reasons for study: To test the effect of variations in the personality & sex of the E upon of performance of M & F Ss in the Draw-A-Person Test.

Hypotheses:

- SOE has a measurable effect upon the drawings produced in the Draw-A-Person Test by M & F Ps.
- The personal characteristics of the E aside form sex have a measurable effect upon the drawings produced by M & F Ps.

Experimenters used: 2 M & 2 F advanced grad students in clinical psyc. Es all differed greatly from one another in terms of phys appearance & personality.

Subjects/Participants used: 40 M & 40 F college students taken from classes in education & psyc at U. of Texas in 1950 (ages 18-48, mean ages of Ms & Fs: 26 & 25).

Experiment design:

- M & F Ss were split randomly & equally btwn M & F Es.
- Es instructed Ss to draw a figure (testing happened individually).
- 12 judges rated:
    - Objective characteristics of the drawing
    - Judged "masculinity" of the figure
    - Intuitive guessing of E's identity
- ANOVA used to analyze data.

Results:

- SOE & personality & appearance of E had no effect on objective characteristics of drawing.
- Ss tended to draw figures of their own sex.

- M Ss drew more “masculine” figures than F Ss – not affected by SOE.

Explanation of results: “…interexaminer variation can be minimized to the point where it is of little importance in the analysis of the indiv Ss’ performance on the Draw-A-Person Test” (148).

Implications: NI

Recommendations: NI

Limitations:
- “Generalization of these findings to the everyday clinical setting where hone is dealing w/ disturbed indivs can be made only w/ great caution” (148).
- Examined only a few of the many possible drawing characteristics.

Problematic methodologies: Androcentric terminology (“Judged ‘masculinity’ of the figure”): probable indication of biased interpretations of drawings.

Author: Harold W. Stevenson
Title: Social Reinforcement with Children as a Function of CA, Sex of E, and Sex of S
Date: 1961
Journal: Journal of Abnormal and Social Psychology
Volume: 63 (1)
Pages: 147-154

Reasons for study: To investigate the effectiveness of social reinforcement in modifying children’s performance in a simple game, paying particular attention to sex of adult providing reinforcement, & sex & age of child being studied.

Hypotheses:
- F Es will have a greater effect than M Es on M & F Ss for Ss ages 3-4.
- Es will have a greater effect on opp-sex Ss than on same-sex Ss for Ss ages 6-7.
- Es will have a greater effect on same-sex Ss than on opp-sex Ss for Ss ages 9-10.

Experimenters used: 6 Ms & 6 Fs involved in Institute of Child Development program. F Es included 2 grad students in psyc, 3 grad students in child development & 1 secretary. M Es included a grad student in psyc, a grad student in child

development, a postdoctoral research associate in child development, & undergrad students in psyc, sociology & industrial ed. Ages 19-25, mean age: 22.8.

Subjects/Participants used: 252 M & 252 F students in preschools & elem schools in Minneapolis & St. Paul – mostly of average intellectual level & from families of average socioeconomic status – selected on basis of availability rather than any other criteria. 1/3 ages 3-4, 1/3 ages 6-7, 1/3 ages 9-10.

Experiment design:

- Each E tested 7 M & 7 F randomly assigned Ss from each age group.
- Es instructed Ss to drop colored marbles one at a time into colored holes going into two bins beside a table,
- For 1st min, E made no response.
- After that, E provided pos reinforcement to S every 30 secs.
- E recorded # of marbles S inserted during each min.

Results:

- F Es had a significantly greater effect than M Es on M & F Ss for Ss ages 3-4.
- No signif SOE effects for other 2 age groups of Ss, although Ps ages 6-7 showed a tendency for better performance w/ opp-sex Ss.
- Social reinforcement delivered by Ms became increasingly effective as Ss' age increased, but no signif changes in general effectiveness of social reinforcement delivered by Fs as a function of Ss' ages was found.
- Ss w/ low base rates had high increments in response & Ss w/ high base rates had lower increments in response.
- There were some E effects (indiv Es elicited dif results than other Es).

Explanation of results:

- Results provide some support for the predictions derived from psychoanalytic theory.
- Tendency for Ss ages 6-7 to do better w/ opp-sex Es- in line w/ Freud's Oedipal theory.
- F Es had bigger effect on Ss ages 3-4 b/c mothers play bigger role in children's lives than fathers at this age.
- No indication exists of what E characteristics produced E effects.

Implications: NI.

Recommendations: NI.

Noted limitations: NI.

Problematic methodologies: This is one of numerous coded studies that use this marble dropping "game." In this case (& in a couple other cases), the instructions from E to S are written out in full. At no time does the E state that the goal is to drop marbles *quickly*, yet rate of dropping is the variable that is monitored while E

provides reinforcement. I think this is a very ethnocentric mode of experimentation…speed may not be of the same intrinsic value for Ss from non-American cultures as it is for Ps raised w/ strictly "U.S." values. Ss may not automatically respond to supportive comments by attempting to go faster.

Author: Gary Alan Fine & Beverly J. Crane
Title: The expectancy effect in anthropological research: an experimental study of riddle collection
Date: 1977
Journal: American Ethnologist
Volume: 4 (3)
Pages: 517-524

Reasons for study: Studies have shown that E expectancy affects research results (& that SOE may also be an important factor), but this phenomenon has not been explored explicitly in anthropology & folklore (e.g. riddling).

Hypotheses:

- Es who are told they should be able to collect many riddles will collect more than Es who are told nothing, while Es who are told riddling is a dying tradition in American will collect fewer.
- There will be a modest positive relationship btwn the # of riddles an informant think they know & the # they are able to tell.

Experimenters used: 6 M & 6 F grad. students who were paid $15 for 3-4 hrs of interviewing.

Subjects/Participants used: 48 M & 48 F undergrad. students of a well-known private univ. Es advertised for Ss in main entry of social sciences building. Ss signed up to be volunteers for a 20-min exp. dealing w/ cultural attitudes. Ss were paid $2 for participation.

Experiment design:

- Es read about riddles. 1/3 read an extra paragraph about popularity of riddles in the U.S., telling them they should be able to collect many. 1/3 read an extra paragraph about riddling as a dying tradition in the U.S. 1/3 read no extra paragraph.
- Ss filled out a questionnaire about personal demographics & approximate # of riddles they thought they knew & whether or not they considered themselves active riddle tellers.
- M & F Ss were divided equally & randomly btwn M & F Es.

- Es interviewed for 15-20 mins, asking Ss to tell riddles.
- Es were allowed to prompt & encourage.
- # of riddles told was analyzed.

Results:

- Ss who considered themselves active riddle tellers. who thought they knew more riddles, & who were young & less educated told more riddles than other Ss.
- Es w/ high expectancy collected the most riddles & Es w/ low expectancy collected the fewest riddles.
- F Es were particularly influenced by high expectancy, & they collected fewer riddles in the neutral condition than in the low expectancy condition.
- M Es collected more riddles in the neutral condition than in the high expectancy condition.
- For Es w/ interviewing experience, the expectancy effect held strongly.
- For Es w/ no interviewing experience, any expectancy (high or low) increased the # of riddles collected.
- There was no interaction btwn SOE & SOS.
- Ss who were interviewed in smaller rooms told more riddles.

Explanation of results: Men may react best an undefined challenge, while women may react better to relatively more structured situations.

Implications:

- "In neutral matters, such as the collection of riddles, interviewers' sex alone does not seem to make a difference" (522).
- People speak more freely in comfortable environments (like in small rooms).

Recommendations:

- "Becoming conscious of expectancy effects may serve as a counterweight to prevent bias" (523).
- Conduct interviews in a casual, personable setting.

Noted limitations: There was no control for the type of riddle collected, which may have varied depending on SOS-SOE interaction.

Problematic methodologies: E demographics (beyond being grad students) not described.

Author: Michael L. Barnes and Robert Rosenthal
Title: Interpersonal Effects of Experimenter Attractiveness, Attire, and Gender

Date: 1985
Journal: Journal of Personality and Social Psychology
Volume: 48 (2)
Pages: 435-446

Reasons for study: Past research has shown that difs among Es can lead to difs in experimental results & that phys. attractiveness & attire affect they way ppl perceive one another.

Hypotheses:
- Es' attributes will affect the Ss' scores on dependent measures.
- Es' attributes will affect their own nonverbal behavior.
- Es' nonverbal behavior will predict the Ss' scores on the dependent measures.

Experimenters used: 6 Fs & 6 Ms.

Subjects/Participants used: 30 F & 30 M Harvard undergrads who were recruited from the dining halls—volunteers willing to donate 1 ½ hrs of their time.

Experiment design:
- Each E interacted w/ 2 F & 2 M Ss.
- Es asked Ss to do a photo-rating task, rating women & men of varying attractiveness as either successful or not.
- Es then administered a vocab. test to Ss.
- Ss left E & were then given an Adjective Check List to fill out according to how they had perceived the E.
- Es & Ss were debriefed (but Es were never told that their phys. attractiveness was a variable).
- 6 F & 3 M undergrads watched E-S interaction on tape. They determined attractiveness level of Es & made judgments about Es' behaviors.
- Results were analyzed w/ 4-way & 5-way ANOVA.

Results:
- F Es were rated higher in self-confidence than M Es (on Adj. Check List).
- Es were rated better on ACL by Ss of opp. sex, an effect which was augmented when Es were better dressed.
- No main effect for Es' phys. attractiveness on any measure.
- M Ss w/ attractive Es & F Ss w/ unattractive Es gave better ratings than other Ss.
- M Ss gave better ratings than F Ss.
- Photos of attractive ppl received higher success ratings from Ss.
- Attractive M Es & unattractive F Es elicited more ratings of success in photo-rating task.
- Es acted more positively toward M Ss.
- Es glanced more often at M Ss than at F Ss.

- Es were rated as more anxious by Ss of same sex.
- Attractive Es behaved less positively than unattractive Es.

Explanation of results:
- Ss' behavior may be to some degree determined by supposedly irrelevant factors.
- Attractiveness may matter less in person than it does in photos.
- It may be that, regardless of one's own sex, ppl treat Ms & Fs differently.

Implications: "These data suggest that no single variable can explain the complexity of behavior in dyadic interactions" (445).

Recommendations: "Gender, physical attractiveness and attire…should be considered explicitly in the planning or analysis of experiments" (445).

Noted limitations: NI.

Problematic methodologies: E demographics not specified.

Author: Wendy J. Nilsen and Scott R. Vrana
Title: Some Touching Situations: The Relationship Between Gender and Contextual Variables in Cardiovascular Responses to Human Touch
Date: 1998
Journal: Annals of Behavioral Medicine
Volume: 20 (4)
Pages: 270-276

Reasons for study: To more systematically investigate the physiological effect of context & gender on touch.

Hypotheses:
- Social touch will cause more arousal than professional touch.
- Social touch will be more susceptible to gender effects.

Experimenters used: 3 Fs & 2 Ms.

Subjects/Participants used: 64 F & 61 M white, native-born Americans (mean age: 19.46) recruited from intro psyc courses, participated for course credit.

Experiment design:

- Ss were met by an F research assistant, who explained & attached heart rate & blood pressure monitoring equipment to Ss.
- Ss sat alone for 2 mins, then M or F E came in and interacted w/ S 3 times (leaving for 2 mins btwn each interaction).
- E used dif. touch condition each of the 3 times (varying the sequence of the 3 conditions w/ each S):
    - Touch wrist for 10 secs to take pulse while looking at a stopwatch
    - No touch, saying the equipment took it automatically
    - Touch wrist for 10 secs w/o explanation while looking at papers, after apologizing for the delay.
- Ss' heart rate & blood pressure was recorded and later analyzed.
- Ss were asked to fill out a questionnaire, and then debriefed.

Results:

- Ss' heart rates in no-touch & pulse touch conditions were lower than base-line.
- Heart rates were higher in social touch condition than in no-touch condition & no-touch condition was more arousing than pulse touch.
- F Es produced greater heart rate deceleration than did M Es.
- For M Es, type of touch didn't matter w/ M Ss, but for F Ss, social touch produced greater heart rates than pulse touch.
- There were steeper declines in heart rate over the 10-sec period of touch for opp-sex pairs.

Explanation of results:

- Changes in heart rate over 10 sec periods may be due to cardiovascular compensatory mechanisms.
- "The automatic arousal evidenced by some participants in the social condition may have been caused by the unspecified and possibly inappropriate function of the touch in that context" (27).
- "…women indicate more comfort w/ same-sex touch than men, but more discomfort w/ opp-sex touch than males" (27).

Implications: "…researchers interested in accounting for variability in cardiovascular response need to understand & attend to the effects of social context in order to explain variability that is often mistakenly attributed to indiv difs or nuisance variance" (27).

Recommendations: "…it is important in all studies of cardiovascular reactivity to report on the characteristics of both the participants & the experimenters" (27).

Noted limitations: NI.

Problematic methodologies:

- Non-whites were excluded from the study, greatly limiting generalizability.
- E demographics not specified.

Author: K. Gijsbers and F. Nicholson
Title: Experimental Pain Thresholds Influenced by Sex of Experimenter
Date: 2005
Journal: Perceptual and Motor Skills
Volume: 101
Pages: 803-807

Reasons for study: To investigate whether a cross-sex effect could be found in an experimental test of pain thresholds rather than pain ratings.

Hypotheses: Ps will show higher pain thresholds when tested by opp-sex Es.

Experimenters: 1 M & 1 F (both 21 yrs old) - dressed to accentuate stereotypical gender characteristics of masculinity & femininity.

Subjects/Participants: 32 M & 32 F volunteers ages 18-49 - all in good health & not using any form of medication.

Experiment design:

- M & F Ps were split equally & randomly btwn M & F Es, who tested them individ'lly.
- Es explained procedure to Ps.
- Ps rated their level of anxiety.
- Using a pressure algometer, Es applied pressure to Ps' upper sternum while Ps' eyes were closed.
- Ps told Es to stop when they reached their pain threshold.
- ANOVA used to analyze data.

Results:

- M Ps had higher pain thresholds than F Ps.
- M Ps had higher pain thresholds w/ F Es than w/ M Es.
- SOE had no effect on F Ps.
- Anxiety level showed no correlation w/ pain thresholds.

Explanation of results: Men are more sensitive to cross-sex testing effects than women (though pressure to the sternum by an M E may have provoked increased anxiety & thus increased pain sensitivity for F Es, even though they reported low levels of anxiety).

Implications: M patients may have a tendency to underreport pain in a clinical setting in the presence of an F physician.

Recommendations: Conduct field studies to test the gender effects found in this study.

Noted limitations: Results may not be applicable to "real life", esp since this study set out to highlight, rather than minimize, sex of administrator effects.

Problematic methodologies:

- Only 1 M & 1 F E used.
- M E wore "a T-shirt & jeans" & F E wore "a skirt, sweater & high heeled shoes" in order to "accentuate stereotypical gender characteristics" - however, it seems to me that the M E was dressed much more casually than the F E…possible confounding variable.

Appendix B

Script: Competent Condition

Throughout the following script, the experimenter remains calm, authoritative, efficient and self-assured. The voice on the tape is steady.

Four participants—three female and one male—enter a room. The experimenter is already there, wearing a white coat and carrying a clipboard.

Experimenter: Please sit down. *S/he gestures toward four chairs. The participants sit.* First of all, I'm going to read your names from this list. *S/he reads names from a list on the clipboard. The participants say,* "Here" *when their names are called.* Okay, we're going to follow the instructions on this tape to run an experiment on heart rate and apprehension. *The experimenter puts a tape in a player and presses "Play."*

Tape: We're going to run a number of trials in which three participants will be randomly assigned to different roles. One participant will be hooked up to an ECG (electrocardiogram), which will monitor his or her heart rate. Two other participants, the "Giver" and the "Taker", will sit opposite each other on either side of a red button. When I say, "Go," the "Giver" and "Taker" are to try to push the red button before the other participant pushes it. When the "Giver" pushes the red button faster than the "Taker", the participant hooked up to the ECG will receive a point. When the "Taker" pushes it faster, the person hooked up to the ECG will lose a point. I will say, "Go" five times per trial. The time that will elapse between each "Go" command will vary. At the end of each trial, the person hooked up to the ECG will receive $20 if he or she

has a positive number of points. He or she will receive nothing with negative points. If he or she already has $20 from a previous trial and then receives negative points, the $20 must be returned to the experimenter. In each trial, the ECG will monitor the participant's heart rate. *The experimenter pauses the tape.*

Experimenter: Is everything clear? *The participants nod.* Okay, then, we'll start with three practice trials. *The experimenter presses "Play."*

Tape: The experimenter has a bag containing four slips of paper designating roles: "Giver", "Taker", "ECG participant", and "Observer." Each of you may now draw a slip of paper from the bag. *The tape pauses, allowing time for the experimenter to hold out the bag to each participant, allowing them to draw slips of paper.* Now the experimenter will hook up the ECG to the ECG participant. *The tape pauses.*

Experimenter: Who's the ECG participant? *A participant raises his/her hand.* Please sit in this chair over here. *The participant does so. The experimenter hooks up the ECG to the participant.*

Tape: Now the "Giver" and "Taker" will sit on either side of the red button.

Experimenter: Please move your chairs over here and take a seat. *The respective participants do so.*

Tape: The "Giver" and "Taker" must keep their hands in their laps between each "Go" command. The experimenter will keep a tally of the ECG participant's points gained and lost throughout each trial. The experimenter will also monitor the ECG's tracking of the ECG participant's heart rate. Now we will begin Practice Trial One. *The tape says,* "Go," *five times with varying amounts of time elapsing between each*

command. The participants reach for the red button, making it so that each time there is a relatively clear "winner." That is the end of Practice Trial One.

Experimenter: The points are *negative/positive* so if this were an actual trial instead of a practice trial, I *would/would not* give you $20.

Tape: Now the experimenter will collect and redistribute the slips of paper for Practice Trial Two. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Practice Trial Two. *The tape says,* "Go,"*five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Practice Trial Two.

Experimenter: The points are *negative/positive* so if this were an actual trial I *would/would not* give you $20 <u>or</u> The points are negative so if this were an actual trial I would take back the $20 I would have given you previously.

Tape: Now the experimenter will collect and redistribute the slips of paper for Practice Trial Three. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Practice Trial Two. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Practice Trial Three.

Experimenter: The points are *negative/positive* so in an actual trial I *would/would not* give you $20 <u>or</u> The points are negative so in an actual trial I would take back the $20 I would have given you previously.

The film cuts to Trial 14.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 14. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 14. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Trial 14.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 15. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 15. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Trial 15.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 16. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 16. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Trial 16.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 17. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 17. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command.*

The "Taker" is faster than the "Giver" on the first three "Go" commands, so the ECG has negative three points so far in this trial. The "Observer" is watching the ECG with an expression of growing anxiety.

Observer: His/her heart rate is getting *really* high! *To the experimenter:* Is s/he okay? *To the ECG participant:* Are you okay? *To the experimenter:* Should we stop the trial?

Experimenter: This has happened before. Yes, we'll stop the trial. *The experimenter stops the tape and removes the ECG from the ECG participant. To the ECG participant:* Are you all right? *The ECG participant nods.* Okay, good, just rest for a minute, and here, have some water. *The experimenter hands the participant a glass of*

water. To all the participants: Okay, I think we'll stop there for now. Thank you all very much for your time.

Script: Incompetent Condition

Throughout the following script, the experimenter stutters periodically, acts confused and inexperienced, and appears unfamiliar with the taped instructions. The voice on the tape is somewhat halting.

Four participants (participants)—three female and one male—enter a room. The experimenter is already there, wearing jeans and a tee shirt, with nothing in hand. **Experimenter:** Please sit down. *S/he gestures toward three chairs. Three participants sit.* Oh…oh, there aren't enough chairs, are there? Hmm, well, sorry, just a second, I'll go get another one. *The experimenter leaves the room and returns with another chair. The participant sits.* Okay, uh…we're going to have to delay the start of the experiment for a minute. There are some problems right now. *S/he fumbles through the drawers of a desk searching for a tape and a clipboard.* Okay, uh, first of all, I'm going to read off your names from this list. *S/he reads the first name on a list*

on a clipboard. No one responds. The experimenter repeats the name. Still no one responds. The experimenter reads the second name on the list. Still no one responds. Oh, this must be the wrong list. Hold on. *S/he fumbles through the drawers again and finds the correct list. S/he reads the names from the list. The participants say,* "Here" *when their names are called.* Okay, we're, um, gonna follow the instructions on this tape to run an experiment on, uh, heart rate and apprehension. *The experimenter drops the tape, picks it up, puts it in the player, and presses "Play."*

Tape: We're going to run a number of trials in which three participants will be randomly assigned to different roles. One participant will be hooked up to an ECG (electrocardiogram), which will monitor his or her heart rate. Two other participants, the "Giver" and the "Taker", will sit opposite each other on either side of a red button. When I say, "Go," the "Giver" and "Taker" are to try to push the red button before the other participant pushes it. When the "Giver" pushes the red button faster than the "Taker", the participant hooked up to the ECG will receive a point. When the "Taker" pushes it faster, the person hooked up to the ECG will lose a point. I will say, "Go" five times per trial. The time that will elapse between each "Go" command will vary. At the end of each trial, the person hooked up to the ECG will receive $20 if he or she has a positive number of points. He or she will receive nothing with negative points. If he or she already has $20 from a previous trial and then receives negative points, the $20 must be returned to the experimenter. In each trial, the ECG will monitor the participant's heart rate. *The experimenter pauses the tape.*

Experimenter: Is everything clear? *The participants nod.* Okay, then, we'll start with three practice trials. *The experimenter presses "Play."*

Tape: The experimenter has a bag containing four slips of paper designating roles: "Giver", "Taker", "ECG participant", and "Observer." Each of you may now draw a slip of paper from the bag. *The tape pauses, allowing time for the experimenter to hold out the bag to each participant, allowing them to draw slips of paper.* Now the experimenter will hook up the ECG to the ECG participant. *The tape pauses.*

Experimenter: Who's the ECG participant? *A participant raises his/her hand.* Please sit in this chair over here. *The participant does so. The experimenter looks confused while hooking up the ECG to the participant.*

Tape: Now the "Giver" and "Taker" will sit on either side of the red button.

Experimenter: Please move your chairs over here and take a seat. *The participants do so.*

Tape: The "Giver" and "Taker" must keep their hands in their laps between each "Go" command. The experimenter will keep a tally of the ECG participant's points gained and lost throughout each trial. The experimenter will also monitor the ECG's tracking of the ECG participant's heart rate. Now we will begin Practice Trial One. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."*

Experimenter *(in the middle of the trial, while tallying points):* Oh, wait a minute, *you're* the Giver and *you're* the Taker…oops…Okay, I'll fix that. *S/he erases something and continues to tally.*

Tape: That is the end of Practice Trial One.

Experimenter: The points are *negative/positive* so if this were an actual trial instead of a practice trial, I *would/would not* give you $20.

Tape: Now the experimenter will collect and redistribute the slips of paper for Practice Trial Two. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Practice Trial Two. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Practice Trial Two.

Experimenter: The points are *negative/positive* so if this were an actual trial I *would/would not* give you $20 <u>or</u> The points are negative so if this were an actual trial I would take back the $20 I would have given you previously.

Tape: Now the experimenter will collect and redistribute the slips of paper for Practice Trial Three. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Practice Trial Three. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Practice Trial Three.

Experimenter: The points are *negative/positive* so in an actual trial I *would/would not* give you $20 <u>or</u> The points are negative so in an actual trial I would take back the $20 I would have given you previously.

The film cuts to Trial 14.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 14. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 14. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Trial 14.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 15. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 15. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red button, making it so that each time there is a relatively clear "winner."* That is the end of Trial 15.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 16. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 16. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command. The participants reach for the red*

button, making it so that each time there is a relatively clear "winner." That is the end of Trial 16.

Experimenter: The points are *negative/positive* so I will give you $20/will not give you $20/will take $20 away from you.

Tape: Now the experimenter will collect and redistribute the slips of paper for Trial 17. *The experimenter does so.* The participants must take their respective places. *The participants do so, with the experimenter setting up the ECG on the appropriate participant.* Now we will begin Trial 17. *The tape says,* "Go," *five times with varying amounts of time elapsing between each command.*

The "Taker" is faster than the "Giver" on the first three "Go" commands, so the ECG has negative three points so far in this trial. The "Observer" is watching the ECG with an expression of growing anxiety.

Observer: His/her heart rate is getting *really* high! *To the experimenter:* Is s/he okay? *To the ECG participant:* Are you okay? *To the experimenter:* Should we stop the trial?

Experimenter *(very distressed)***:** This has never happened before. Um, yes, we'll stop the trial. *The experimenter stops the tape and removes the ECG from the participant. To the ECG participant:* Are you all right? Should I go get help?

ECG participant: That's Okay, I'm fine.

E: Okay, um…good. *To all participants:* Okay, I think we'll stop there for now. Thanks for all your time.

Appendix C

Consent Form

I state that I am 18 years of age or older and agree to participate in a program of research being conducted by Professor Morawski and Thesis Tutorial student Brittany Morse of the Wesleyan University Psychology Department. The focus of this research is on the effects of participant feedback on experimenters' subsequent performance/behavior. The risk involved is no greater than that found in everyday life. A benefit of participation includes a possible increased understanding of the effects of participants' feedback on experimenters' performance/behavior, as participants may—if they so choose—have access to the results of the study.

I understand that all of my responses will be held in strict confidence and will not be identified in any publication of the results. Specifically, I understand that the data collected from this research will be stored in Room 303 of Judd Hall in locked cabinets and that only Professor Morawski and Brittany Morse will have access to the data. The data will be coded so as not to identify me by name and will be destroyed after five years.

I understand that in return for serving in this experiment, I will receive one hour of credit toward the Research Participation requirement in Psychology 105. I further understand that participation in the research is voluntary, that I may ask questions, and that I am free to withdraw from the experiment at any time. I understand that if I choose to withdraw during the experiment, I will still receive full credit. I further understand that I will be given a copy of this consent form to keep for my own records. Finally, I understand that if I have any comments, questions, or concerns following the experiment, I may contact Brittany Morse by telephone (617-697-4732), by e-mail (brittany.morse@gmail.com), in person (Room 303 of Judd Hall) or I may contact Professor Jill Morawski by phone (860-685-2344), by e-mail (jmorawski@wesleyan.edu) or in person (Room 317 of Judd Hall). I may also bring complaints about the experiment to Ruth Striegel-Moore, Chair of the Wesleyan Psychology Department (860-685-2328).

Name of participant (print clearly):

Signature of participant: _____ Date: _____

EXPERIMENTER COPY

Consent Form

I state that I am 18 years of age or older and agree to participate in a program of research being conducted by Professor Morawski and Thesis Tutorial student Brittany Morse of the Wesleyan University Psychology Department. The focus of this research is on the effects of participant feedback on experimenters' subsequent performance/behavior. The risk involved is no greater than that found in everyday life. A benefit of participation includes a possible increased understanding of the effects of participants' feedback on experimenters' performance/behavior, as participants may—if they so choose—have access to the results of the study.

I understand that all of my responses will be held in strict confidence and will not be identified in any publication of the results. Specifically, I understand that the data collected from this research will be stored in Room 303 of Judd Hall in locked cabinets and that only Professor Morawski and Brittany Morse will have access to the data. The data will be coded so as not to identify me by name and will be destroyed after five years.

I understand that in return for serving in this experiment, I will receive one hour of credit toward the Research Participation requirement in Psychology 105. I further understand that participation in the research is voluntary, that I may ask questions, and that I am free to withdraw from the experiment at any time. I understand that if I choose to withdraw during the experiment, I will still receive full credit. I further understand that I will be given a copy of this consent form to keep for my own records. Finally, I understand that if I have any comments, questions, or concerns following the experiment, I may contact Brittany Morse by telephone (617-697-4732), by e-mail (brittany.morse@gmail.com), in person (Room 303 of Judd Hall) or I may contact Professor Jill Morawski by phone (860-685-2344), by e-mail (jmorawski@wesleyan.edu) or in person (Room 317 of Judd Hall). I may also bring complaints about the experiment to Ruth Striegel-Moore, Chair of the Wesleyan Psychology Department (860-685-2328).

Name of participant (print clearly):

Signature of participant: _____ Date: _____

PARTICIPANT COPY

Appendix D

Please provide your:

Sex/Gender: _____

Class year: _____

Race/Ethnicity (optional): _____

Major(s) (if declared): _____

Please provide your ratings of the experimenter in the film:

Evaluative:

good _____:_____:_____:_____:_____:_____:_____ bad

successful _____:_____:_____:_____:_____:_____:_____ unsuccessful

positive _____:_____:_____:_____:_____:_____:_____ negative

wise _____:_____:_____:_____:_____:_____:_____ foolish

attractive _____:_____:_____:_____:_____:_____:_____ unattractive

Potency:

hard _____:_____:_____:_____:_____:_____:_____ soft

masculine _____:_____:_____:_____:_____:_____:_____ feminine

strong _____:_____:_____:_____:_____:_____:_____ weak

severe _____:_____:_____:_____:_____:_____:_____ lenient

serious _____:_____:_____:_____:_____:_____:_____ humorous

Activity:

fast _____:_____:_____:_____:_____:_____:_____ slow

active _____:_____:_____:_____:_____:_____:_____ passive

excitable _____:_____:_____:_____:_____:_____:_____ calm

impulsive _____:_____:_____:_____:_____:_____:_____ deliberate

simple _____:_____:_____:_____:_____:_____:_____ complex

Competency:

confident _____:_____:_____:_____:_____:_____:_____ nervous

professional _____:_____:_____:_____:_____:_____:_____ novice

efficient _____:_____:_____:_____:_____:_____:_____ inefficient

experienced
inexperienced
with the
instruments _____:_____:_____:_____:_____:_____:_____ with the
instruments

competent _____:_____:_____:_____:_____:_____:_____ incompetent

Appendix E

Dear Participant,

Thank you for your participation! We would now like to debrief you on the true nature of our research.

We told you that the study was on the effects of participants' feedback on experimenters' subsequent performance/behavior. Actually, the study is on participants' perceptions/judgments of male and female experimenters behaving either competently or incompetently. We deceived you, our participants, regarding the purpose of the study so as not to influence your expressed perceptions/judgments of the videotaped experimenters.

We are conducting this study as a modified replication of a study conducted by Beth Stearns Piacente in 1974 in order to determine the degree to which her findings remain prevalent today.

Some of Piacente's findings include:

- male Es (experimenters) were perceived as *better* and more *masculine* than female Es,
- female Es were perceived as more *excitable* and *attractive* than male Es,
- male (but not female) Ps (participants) rated male Es as more *efficient* than female Es.
- competent female and male Es were rated fairly equally,
- incompetent female Es were rated more negatively than incompetent male Es,
- and incompetent female Es were perceived as more feminine than competent female Es when Ps were run by a male administrator but not when they were run by a female administrator.

To read more about Piacente's findings, you may access the following reports:

Piacente, B. S. (1974). Women as Experimenters. *American Psychologist, 29 (7),* 526-529.

Piacente, B.S. (1974). Evaluation of the Performance of Experimenters as a Function of Their Sex and Competence. *Journal of Applied Social Psychology, 4 (4), 321-329.*

We will keep you informed of the results of our data analysis unless you request otherwise. If you have any questions or if you would like to discuss the effects and/or the necessity of the deception used in this study, please contact either Brittany Morse (617-697-4732, brittany.morse@gmail.com) or Professor Jill Morawski (860-685-2344, jmorawski@wesleyan.edu).

Thank you again for your participation!

References

Abrams, D., Viki, G. T., Masser, B., & Gerd, B. (2003). Perceptions of stranger and acquaintance rape: The role of benevolent and hostile sexism in victim blame and rape proclivity. *Journal of Personality and Social Psychology, 84*(1), 111-125.

Astin, H. S. (1973). Report of the Task Force on the Status of Women in Psychology. *American Psychologist, 28*(7), 611-616.

Barreto, M., Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology, 35*(5), 633-642.

Basow, S. A., & Silberg, N. T. (1987). Student Evaluations of College Professors: Are Female and Male Professors Rated Differently? *Journal of Educational Psychology, 79*(3), 308-314.

Blau (2000). Gender Differences in Pay. *Journal of Economic Perspectives, 14*(4), 75-99.

Busk-Jensen, L. (1985). The woman, women, femininity: Considerations to the fundamental concepts of women's studies. *Psyke & Logos, 6*(1), 7-24.

Carringer, D., & Wilson, C. S. (1974). The effects of sex, socioeconomic class, experimenter race, and kind of verbal reinforcement on the performance of black children. *Journal of Negro Education, 43*(2), 212-220.

Condor, S. (1991). Sexism in psychological research: A brief note. *Feminism & Psychology, 1*(3), 430-434.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology, 93*(5), 764-779.

Denmark, F., Russo, N. F., Frieze, I. H., & Sechzer, J. A. (1988). Guidelines for Avoiding Sexism in Psychological Research: A Report of the Ad Hoc Committee on Nonsexist Research. *American Psychologist, 43*(7), 582-585.

Diamond, E. E. (1976). Minimizing sex bias in testing. *Measurement & Evaluation in Guidance, 91*(1), 28-34.

Dutton, D. G., & Yee, P. (1974). The effect of subject liberalism, anonymity, and race of experimenter on subjects' ratings of oriental and white photographs. *Canadian Journal of Behavioural Science, 6*(4), 332-341.

Eagly, A. H. (1978). Sex Differences in Influenceability. *Psychological Bulletin, 85*(1), 86-116.

Eagly, A. H., Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*(3), 283-308.

Eizner Favreau, O. (1977). Sex bias in psychological research. *Canadian Psychological Review, 18*(1), 56-65.

Feingold, A. (1988). Cognitive Gender Differences Are Disappearing. *American Psychologist, 43*(2), 95-103.

Fox, R. L., Lawless, J. L., & Feeley, C. (2001). Gender and the decision to run for office. *Legislative Studies Quarterly, 16*(3), 411-435.

Gannon, L., Luchetta, T., Rhodes, K., Pardie, L., & Segrist, D. (1992). Sex Bias in Psychological Research: Progress or Complacency? *American Psychologist, 47*(3), 389-396.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist, 56*(2), 109-118.

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of Women Quarterly, 21*(1), 119-135.

Gordon, M. E., Slade, L. A, & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. *Academy of Management Review, 11*(1), 191-207.

Grady, K. E. (1981). Sex bias in research design. *Psychology of Women Quarterly, 5*(4), 628-636.

Goldberg, P. (1968). Are women prejudiced against women? *Transaction, 5,* 316–322.

Guthrie, R. V. (1976). *Even the Rat Was White: A Historical View of Psychology.* New York: Harper & Row.

Guthrie, R. V. (1998). *Even the Rat Was White: A Historical View of Psychology* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Gray, V. A. (1977). The Image of Women in Psychology Textbooks. *Canadian Psychological Review, 18*(1), 46-55.

Hare-Mustin, R. T., & Marecek, J. (1990). *Making a difference: Psychology and the construction of gender.* New Haven: Yale University Press.

Harris, S. (1971). Influence of subject and experimenter sex in psychological research. *Journal of Consulting and Clinical Psychology, 37*(2), 291-294.

Heilman, M. E. (2001). Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder. *Journal of Social Issues, 57*(4), 657-674.

Hojat, M., Gonnella, J. S., Mangione, S., Nasca, T. J., Veloski, J. J., Erdmann, J. B., Callahan, C. A., & Magee, M. (2002). Empathy in medical students as related to academic performance, clinical competence and gender. Medical Education, 36(6), 522-527.

Jung, J. (1969). Current practices and problems in the use of college students for psychological research. *Canadian Psychologist, 10*(3), 280-290.

Kasof, J. (1993). Sex Bias in the Naming of Stimulus Persons. *Psychological Bulletin, 113*(1),140-163.

Levine (2008). Wallowing in Sex: The New Sexual Culture in 1970s American Television. *Journal of Popular Culture, 41*(1), 173-175.

McHugh, M. C., Koeske, R. D., & Frieze, I. H. (1986). Issues to consider in conducting nonsexist psychological research: A guide for researchers. *American Psychologist, 41*(8), 879-890.

Mead, M. (1978). The relationship between research by women and women's experiential roles. *Psychology of Women Quarterly, 2*(4), 363-365.

Mednick, M. T. S., & Weisman, H. J. (1975). The psychology of women—Selected topics. *Annual Review of Psychology, 26*, 1-18.

Moon, M., & Hoffman, C. D. (2000). References on Men and Women in Psychology (1887 1997): PsycINFO as an Archival Research Tool. *Psychology of Men and Masculinity, 1*(1), 16-20.

Morawski, J. G. (1997). White Experimenters, White Blood, and Other White Conditions: Locating the Psychologist's Race. In M. Fine, L. Weis, L. C. Powell, & M. Wong, (ed.), *Off White: Reading on Race, Power, and Society* (pp. 13-28). New York, NY: Routledge.

Morawski, J. G. (2006). Epistemological Dizziness: Subject and Experimenter in the Cold War Psychology Laboratory. Unpublished manuscript: Wesleyan University, Middletown, CT.

O'Neil, J. M. (1981). Patterns of gender role conflict and strain: Sexism and fear of femininity in men's lives. *Personnel & Guidance Journal, 60*(4), 203-210.

Paludi, M. A., & Bauer, W. D. (1983). Goldberg revisited: What's in an author's name? *Behavioral Science, 9*(3), 387-390.

Penner., L. A., Hawkins, H. L., Dertke, M. C., Spector, P., & Stone, A. The bumbling experimenter: Obedience as a function of the experimenter competence. *Memory and Cognition, 1*, 241-245.

Piacente, B. S. (1974). Status of women in the psychological community in the Southeast: Women as experimenters. *American Psychologist, 29*(7), 526-529.

Piacente, B. S., Penner, L. A., Hawkins, H. L., & Cohen, S. L. (1974). Evaluation of the performance of experimenters as a function of their sex and competence. *Journal of Applied Social Psychology, 4*(4), 321-329.

Piliavin, J. A., & Martin, R. R. (1978). The effects of the sex composition of groups on style of social interaction. *Sex Roles, 4*(2), 281-296.

Powers, R. S., & Reiser, C. (2005). Gender and self-perceptions of social power. *Social Behavior and Personality, 33*(6), 553-568.

Rosenberg (1982). *Beyond Separate Spheres: Intellectual Roots of Modern Feminism.* New Haven: Yale University Press.

Rosenthal, R. (1966). Biosocial Attributes. In Rosenthal, R., *Experimenter Effects in Behavioral Research* (pp. 38-61). New York, NY: Meredith Publishing Company.

Rossiter (1982). *Women Scientists in America: Struggles and Strategies to 1940.* Baltimore: The Johns Hopkins University Press.

Rumenik, D. K., Capasso, D. R., & Hendrick, C. (1977). Experimenter sex effects in behavioral research. *Psychological Bulletin, 84*(5), 852-877.

Russo, N. F. (1983). Psychology's Foremothers: Their Achievements in Context. In A. N. O'Connell, & N. F. Russo, (ed.), *Models of Achievement: Reflections of Eminent Women in Psychology* (pp. 9-24). New York, NY: Columbia University Press.

Scarborough, E., &Furumoto , L. (1987). *Untold Lives: The First Generation of American Women Psychologists.* New York: Columbia University Press.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*(3), 515-530.

Sherif OL HQ, C. W. (1994). Bias in psychology. Herrmann, Anne C (Ed); Stewart, Abigail J (Ed). (1994). Theorizing feminism: Parallel trends in the humanities and social sciences.

Sibley, C. G., & Wilson, M. C. (2004). Differentiating Hostile and Benevolent Sexist Attitudes Toward Positive and Negative Sexual Female Subtypes. *Sex Roles, 51*(11 12), 687-696.

Silverman, I. (1974). The Experimenter: A (Still) Neglected Stimulus Object. *Canadian Psychologist, 15*(3), 258-270.

Smart, R. G. (1966). Subject selection bias in psychological research. *Canadian Psychologist, 7,* 115-121.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay Versus John McKay: Do Gender Stereotypes Bias Evaluations? *Psychological Bulletin, 105*(3), 409-429.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist, 34*(11), 1085-1094.

Unger, R. K. (2001). Women as subjects, actors, and agents in the history of psychology. In Unger, Rhoda K (Ed), *Handbook of the Psychology of Women and Gender* (pp. 3-16). Hoboken, NJ: John Wiley & Sons Inc.

Unger, R. K. (2007). Afterword: From inside and out: Reflecting on a feminist politics of gender in psychology. *Feminism & Psychology, 17*(4), 487-494.

Weisstein (1968). *Kinder, Kirche, Kuche as Scientific Law: Psychology Constructs the Female.* Boston: New England Free Press.

Wojciszke, B., & Mikiewicz, A. (2003). How parents stress morality and competence in socialization of boys and girls. *Przeglad Psychologiczny, 46*(4), 425-439.

Coded studies

Abramson, P. R., & Handschumacher, I. W. (1978). Experimenter effects on responses to double-entendre words. *Journal of Personality Assessment, 42*(6), 592-596.

Adair, J. G. (1972). Demand characteristics or conformity?: Suspiciousness of deception and experimenter bias in conformity research. *Canadian Journal of Behavioural Science, 4*(3), 238-248.

Alden, P., & Benton, A. L. (1951). Relationship of sex of examiner to incidence of Rorschach responses with sexual content. *Journal of Projective Techniques 15 1951, 231-234*.

Allen, J. L., Sipes, S. J., & Sipes, G. P. (1977). Sex of the experimenter as a variable in the autokinetic illusion. *Bulletin of the Psychonomic Society, 9*(6), 397-398.

Allen, S. A., Spear, P. S., & Johnson, J. (1969). Experimenter role effects on children's task performance and perception. *Child Development 40(1) 1969, 1-9 Blackwell Publishing*.

Archer, E., Cejka, J. A., & Thompson, C. P. (1961). Serial-trigram learning as a function of differential meaningfulness and sex of subjects and experimenters. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 15*(3), 148-153.

Atkins, C., Morse, C., & Zweigenhaft, R. (1978). The stereotype and recognition of female athletes. *Journal of Psychology: Interdisciplinary and Applied, 100*(1), 27-31.

Axelrod, H. S., Cowen, E. L., & Heilizer, F. (1956). The correlates of manifest anxiety in stylus maze learning. *Journal of Experimental Psychology, 51*(2), 131-138.

Barnes, M. L., & Rosenthal, R. (1985). Interpersonal effects of experimenter attractiveness, attire, and gender. *Journal of Personality and Social Psychology, 48*(2), 435-446.

Baucom, D. H., & Danker-Brown, P. (1979). Influences of sex roles on the development of learned helplessness. *Journal of Consulting and Clinical Psychology, 47*(5), 928-936.

Beck, S. J., & Ollendick, T. H. (1976). Personal space, sex of experimenter, and locus of control in normal and delinquent adolescents. *Psychological Reports, 38*(2), 383-387.

Berkowitz, L., Klanderman, S., & Harris, R. (1964). Effects of experimenter awareness and sex of subject and experimenter on reactions to dependency relationship. *Sociometry 27(3) 1964, 327-337*.

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707-724.

Binning, J. F., Goldstein, M. A., Garcia, M. F., & Scattaregia, J. H. (1988). Effects of preinterview impressions on questioning strategies in same- and opposite-sex employment interviews. *Journal of Applied Psychology, 73*(1), 30-37.

Bleda, P. R., Bleda, S. E., & White, L. A. (1976). Bystander reactions to a petty crime. *Criminology: An Interdisciplinary Journal, 14*(2), 287-290.

Borstelmann, L. (1961). Sex of experimenter and sex-typed behavior of young children. *Child Development 32 1961, 519-524 Blackwell Publishing*.

Boutcher, S. H., Fleischer-Curtian, L. A., & Gines, S. D. (1988). The effects of self-presentation on perceived exertion. *Journal of Sport & Exercise Psychology, 10*(3), 270-280.

Brekke, B., & Williams, J. D. (1973). Conservation and sex. *Perceptual and Motor Skills, . 37*(1), 14.

Breyer, N. L., May, J. G., & Gable, R. K. (1972). Effects of race and sex of experimenter and subject on responsiveness during nonreinforcement and reinforcement conditions. *Psychological Reports, . 31*(2), 515-524.

Brockner, J., Pressman, B., Cabitt, J., & Moran, P. (1982). Nonverbal intimacy, sex, and compliance: A field study. *Journal of Nonverbal Behavior, 6*(4), 253-258.

Bryson, S. E., & Pilon, D. J. (1984). Sex differences in depression and the method of administering the Beck Depression Inventory. *Journal of Clinical Psychology, 40*(2), 529-534.

Burton, R. V., Allinsmith, W., & Maccoby, E. E. (1966). Resistance to temptation in relation to sex of child, sex of experimenter, and withdrawal of attention. *Journal of Personality and Social Psychology, 3*(3), 253-258.

Carter, L. E., McNeil, D. W., Vowles, K. E., Sorrell, J. T., Turk, C. L., Ries, B. J., et al. (2002). Effects of emotion on pain reports, tolerance and physiology. *Pain Research & Management, 7*(1), 21-30.

Cason, H., & Cason, E. (1925). Association tendencies and learning ability. *Journal of Experimental Psychology, 8*(3), 167-189.

Cieutat, V. J. (1962). Sex differences and reinforcement in the conditioning and extinction of conversational behavior. *Psychological Reports 10(2) 1962, 467-474 Psychological Reports*.

Cieutat, V. J. (1965). Examiner differences with the Stanford-Binet IQ. *Perceptual and Motor Skills 20(1) 1965, 317-318 Perceptual & Motor Skills*.

Clark, R. A. (1952). The projective measurement of experimentally induced levels of sexual motivation. *Journal of Experimental Psychology, 44*(6), 391-399.

Clark, S., & Lane, M. K. (1978). Women's behavioral manifestations of traditionalist and liberated role concepts. *Journal of Psychology: Interdisciplinary and Applied, 98*(1), 81-89.

Cull, J. G. (1976). The relationship between sex role and modification of judgments. *Journal of Psychology: Interdisciplinary and Applied, 93*(2), 313-317.

Cutler, A., & Scott, D. R. (1990). Speaker sex and perceived apportionment of talk. *Applied Psycholinguistics, 11*(3), 253-272.

Datta, L. E., & Drake, A. K. (1968). Examiner Sex and Sexual Differentiation in Preschool Children's Figure Drawings. *Journal of Projective Techniques & Personality Assessment 32(4) 1968, 397-399 Lawrence Erlbaum*.

Deus, V., & Jokic-Begic, N. (2006). Personal space in schizophrenic patients. *Psychiatria Danubina, 18*(3-4), 150-158.

Dixit, R. (1969). Sex of experimenter as a variable in verbal conditioning. *Indian Journal of Experimental Psychology, . 3*(1), 11-15.

Doll, P. A., Fagot, H. J., & Himbert, J. D. (1971). Experimenter effect on sex-role preference among black and white lower-class male children. *Psychological Reports, . 29*(3, Pt. 2), 1295-1301.

Doll, P. A., Fagot, H. J., & Himbert, J. D. (1973). Examiner effect on sex-role preference among black and white lower-class female children. *Psychological Reports, . 32*(2), 427-434.

Dusek, J. B. (1971). Experimenter bias in performance in children at a simple motor task. *Developmental Psychology, 4*(1, Pt.1), 55-62.

Dweck, C. S., & Bush, E. S. (1976). Sex differences in learned helplessness: I. Differential debilitation with peer and adult evaluators. *Developmental Psychology, 12*(2), 147-156.

Edwards, D. J. (1977). Perception of crowding and personal space as a function of locus of control, arousal seeking, sex of experimenter, and sex of subject. *Journal of Psychology: Interdisciplinary and Applied, 95*(2), 223-229.

Etaugh, C., Houtler, B. D., & Ptasnik, P. (1988). Evaluating competence of women and men: Effects of experimenter gender and group gender composition. *Psychology of Women Quarterly, 12*(2), 191-200.

Etaugh, C. F. (1969). Experimenter and subject variables in verbal conditioning. *Psychological Reports 25(2) 1969, 575-580 Psychological Reports.*

Feine, J. S., Bushnell, M., Miron, D., & Duncan, G. H. (1991). Sex differences in the perception of noxious heat stimuli. *Pain, 44*(3), 255-262.

Fellows, B. J. (1979). The British use of the Barber Suggestibility Scale: Norms, psychometric properties and the effects of the sex of the subject and of the experimenter. *British Journal of Psychology, 70*(4), 547-557.

Ferguson, D. C., & Buss, A. H. (1960). Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *Journal of Consulting Psychology, 24*(4), 324-327.

Fine, G. A., & Crane, B. J. (1977). The expectancy effect in anthropological research: An experimental study of riddle collection. *American Ethnologist, 4*(3), 517-524.

Fischer, G. J. (1977). Probability learning and attitude toward women as a function of monetary risk, gain, and sex. *Bulletin of the Psychonomic Society, 9*(3), 201-203.

Fisher-Thompson, D., & Burke, T. A. (1998). Experimenter influences and children's cross-gender behavior. *Sex Roles, 39*(9-10), 669-684.

Freides, D. (1975). Information complexity and cross-modal functions. *British Journal of Psychology, 66*(3), 283-287.

Galbraith, G. G., & Crow, C. (1976). Retrospective parental ratings and free associative sexual responsivity in male and female college students. *Psychological Reports, 38*(3, Pt 1), 759-765.

Gall, M., & Mendelsohn, G. A. (1967). Effects of Facilitating Techniques and Subject-Experimenter Interaction on Creative Problem Solving. *Journal of Personality and Social Psychology, 5*(2), 211-216.

Galla, J. P., Frison, J. D., Jeffrey, L. R., & Gaer, E. P. (1981). Effect of experimenter's gender on responses to a sex-role attitude questionnaire. *Psychological Reports, 49*(3), 935-940.

Gallagher, P. L., & Rollins, J. (1986). Increasing treatment efficiency through effective intake interviews. *Journal of Drug Education, 16*(1), 35-41.

Gardner, W. I., & Kaufman, M. E. (1968). Verbal conditioning in noninstitutional mildly retarded adolescents as a function of sex of subject and sex of experimenter. *Psychological Reports 23(1) 1968, 207-212 Psychological Reports.*

Garfield, S. L., Blek, L., & Melker, F. (1952). The influence of method of administration and sex differences on selected aspects of TAT stories. *Journal of Consulting Psychology, 16*(2), 140-144.

Garrett, C. S., & Cunningham, D. J. (1974). Effects of vicarious consequences and model and experimenter sex on imitative behavior in first-grade children. *Journal of Educational Psychology, 66*(6), 940-947.

Gijsbers, K., & Nicholson, F. (2005). Experimental pain thresholds influenced by sex of experimenter. *Perceptual and Motor Skills, 101*(3), 803-807.

Gold, D. (1969). Effect of the experimenter in human eyelid conditioning. *Psychonomic Science, 17*(4), 232-233.

Gralton, M. A., Hayes, Y. A., & Richardson, J. T. (1979). Introversion-extraversion and mental imagery. *Journal of Mental Imagery, 3*(1-2), 1-10.

Green, R. J., Sandall, J. C., & Phelps, C. (2005). Effect of Experimenter Attire and Sex on Participant Productivity. *Social Behavior and Personality, 33*(2), 125-132.

Halperin, M. S. (1977). Sex differences in children's responses to adult pressure for achievement. *Journal of Educational Psychology, 69*(2), 96-100.

Harney, D. M., & Parker, R. (1972). Effects of social reinforcement, subject sex, and experimenter sex on children's motor performance. *Research Quarterly Vol 43(2) May 1972, 187-196*.

Harris, S., & Masling, J. (1970). Examiner sex, subject sex, and Rorschach productivity. *Journal of Consulting and Clinical Psychology, 34*(1), 60-63.

Hartup, W. W., & Himeno, Y. (1959). Social isolation vs. interaction with adults in relation to aggression in preschool children. *Journal of Abnormal & Social Psychology, 59*(1), 17-22.

Herzog, E. W., Enright, M., Luria, Z., & Rubin, J. Z. (1982). Do gender labels yield sex differences in performance, or is label a fable? *Developmental Psychology, 18*(3), 424-430.

Hewitt, J., & Feltham, D. (1982). Differential reaction to touch by men and women. *Perceptual and Motor Skills, 55*(3, Pt 2), 1291-1294.

Hill, K. T. (1967). Social Reinforcement as a Function of Test Anxiety and Success-Failure Experiences. *Child Development 38(3) 1967, 723-737 Blackwell Publishing*.

Holtzman, W. H. (1952). The examiner as a variable in the Draw-A-Person Test. *Journal of Consulting Psychology, 16*(2), 145-148.

Jemelka, J. R., & Downs, A. (1991). The social support construct in an experimental context: A reexamination. *Psychology: A Journal of Human Behavior, 28*(3-4), 35-46.

Jensen, K. W. (1985). Sex of the administrator as a variable affecting drawing a person. *Perceptual and Motor Skills, 60*(1), 72-74.

Johnson, R. W. (1970). Subject performance as affected by experimenter expectancy, sex of experimenter, and verbal reinforcement. *Canadian Journal of Behavioural Science, 2*(1), 60-66.

Johnson, R. W., & Adair, J. G. (1972). Experimenter expectancy vs systematic recording error under automated and nonautomated stimulus presentation. *Journal of Experimental Research in Personality Vol 6(1) Mar 1972, 88-94*.

Jokerst, M., Gatto, M., Fazio, R., Gianaros, P. J., Stern, R. M., & Koch, K. L. (1999). Effects of gender of subjects and experimenter on susceptibility to motion sickness. *Aviation, Space, and Environmental Medicine, 70*(10, Sect 1), 962-965.

Kalin, L. R., & Schuldt, W. (1991). Effects of self-awareness on self-disclosure. *Psychological Reports Vol 69(1) Aug 1991, 289-290*.

Kallai, I., Barke, A., & Voss, U. (2004). The effects of experimenter characteristics on pain reports in women and men. *Pain, 112*(1-2), 142-147.

Katz, P. A., & Walsh, P. (1991). Modification of children's gender-stereotyped behavior. *Child Development, 62*(2), 338-351.

Kelley, K. (1985a). Sex, sex guilt, and authoritarianism: Differences in responses to explicit heterosexual and masturbatory slides. *Journal of Sex Research, 21*(1), 68-85.

Kelley, K. (1985b). Sexuality and hostility of authoritarians. *The High School Journal Vol 68(3) Feb-Mar 1985, 173-176*.

Kennedy, J. J., Cook, P. A., & Brewer, R. R. (1970). The effects of three selected experimenter variables in verbal conditioning research. *Journal of Social Psychology 81(2) 1970, 167-175 Heldref Publications.*

Lennon, R., Eisenberg, N., & Carroll, J. (1983). The assessment of empathy in early childhood. *Journal of Applied Developmental Psychology, 4*(3), 295-302.

Leventhal, G., & Herbert, H. (1990). Effects of subjects' sex and sex-role attitudes on real and perceived performance using gender-related stimuli with traditional and nontraditional female experimenters. *Psychological Reports, 66*(1), 259-266.

Levine, F. M., & de Simone, L. L. (1991). The effects of experimenter gender on pain report in male and female subjects. *Pain, 44*(1), 69-72.

Littig, L. W., & Branch, D. (1993). Exposure frequencies and sensation-seeking: No novelty effect but an unexpected experimenter^subject sex interaction. *Social Behavior and Personality, 21*(1), 25-31.

Lundstrom, J. N., & Olsson, M. J. (2005). Subthreshold amounts of social odorant affect mood, but not behavior, in heterosexual women when tested by a male, but not a female, experimenter. *Biological Psychology, 70*(3), 197-204.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin Vol 28(9) Sep 2002, 1183-1193*.

Masling, J., & Harris, S. (1969). Sexual aspects of TAT administration. *Journal of Consulting and Clinical Psychology, 33*(2), 166-169.

Mazzocco, A., Graus, G., & Sonino, M. (1977). The greater regard of the opposite sex. *Italian Journal of Psychology, 4*(1), 113-119.

McMahan, E. D. (1976). The role of the experimenter in observed sex differences in conservation acquisition. *Journal of Psychology: Interdisciplinary and Applied, 94*(2), 205-206.

Motanelli, D. S., & Hill, K. T. (1969). Children's achievement expectations and performance as a function of two consecutive reinforcement experiences, sex of subject, and sex of experimenter. *Journal of Personality and Social Psychology, 13*(2), 115-128.

Nilsen, W. J., & Vrana, S. R. (1998). Some touching situations: The relationship between gender and contextual variables in cardiovascular responses to human touch. *Annals of Behavioral Medicine, 20*(4), 270-276.

O'Sullivan, M., Gilner, F. H., & Krinski, R. (1973). The influence of sex of experimenter onmodeling in the reduction of fear. *Behavior Therapy, . 4*(4), 535-542.

Odom, R. D. (1966). Children's probability learning as a function of the cross-sex effect. *Psychonomic Science, 4*(8), 305-306.

Ogawa, J., & Oakes, W. F. (1965). Sex of experimenter and manifest anxiety as related to verbal conditioning. *Journal of Personality 33(4) 1965, 553-569 Blackwell Publishing*.

Otto, M. W., & Dougher, M. J. (1985). Sex differences and personality factors in responsivity to pain. *Perceptual and Motor Skills, 61*(2), 383-390.

Paludi, M. A., & Bauer, W. D. (1979). Impact of sex of experimenter on the Draw-A-Person Test. *Perceptual and Motor Skills, 49*(2), 456-458.

Parish, C. R., & Wheatley, G. H. (1973). Sex of experimenter as a variable in Piagetian conservation assessment. *Journal of Genetic Psychology, . 123*(2), 273-278.

Pedersen, D. M., Shinedling, M. M., & Johnson, D. L. (1968). Effects of sex of examiner and subject on children's quantitative test performance. *Journal of Personality and Social Psychology, 10*(3), 251-254.

Peplin, C. M., & Larsen, J. D. (1989). Field dependence: Experimenters' expectancy, sex difference or no difference? *Perceptual and Motor Skills, 68*(2), 355-358.

Piacente, B. S. (1974). Status of women in the psychological community in the Southeast: Women as experimenters. *American Psychologist, 29*(7), 526-529.

Piacente, B. S., Penner, L. A., Hawkins, H. L., & Cohen, S. L. (1974). Evaluation of the performance of experimenters as a function of their sex and competence. *Journal of Applied Social Psychology, 4*(4), 321-329.

Powell, J. L., Meil, W., Patterson, M. L., Chouinard, E. F., & et al. (1994). Effects of timing of touch on compliance to a request. *Journal of Social Behavior & Personality, 9*(1), 153-162.

Quereshi, M. (1968). Intelligence Test Scores as a Function of Sex of Experimenter and Sex of Subject. *Journal of Psychology: Interdisciplinary and Applied 69(2) 1968, 277-284 Heldref Publications.*

Rekers, G. A., & Rudy, J. P. (1978). Differentiation of childhood body gestures. *Perceptual and Motor Skills, 46*(3, Pt 1), 839-845.

Remmers, H., Cutler, M., & Jones, P. (1940). Waking suggestibility in children--general or specific? *Journal of Genetic Psychology 56 1940, 87-93 Heldref Publications*.

Reysen, S., & Ganz, E. (2006). Gender differences in helping in six U.S. cities. *North American Journal of Psychology, 8*(1), 63-67.

Rikli, R. (1974). Effects of experimenter expectancy set and experimenter sex upon grip strength and hand steadiness scores. *Research Quarterly Vol 45(4) Dec 1974, 416-423*.

Rikli, R. (1976). Physical performance scores as a function of experimenter sex and experimenter bias. *Research Quarterly Vol 47(4) Dec 1976, 776-782*.

Robertson, K., & Murachver, T. (2003). Children's speech accommodation to gendered language styles. *Journal of Language and Social Psychology, 22*(3), 321-333.

Rosenbluh, E. S. (1972). Interaction of subject and experimenter sex with the body percept in verbal concept identification. *Journal of Abnormal Psychology, 80*(2), 149-156.

Rowley, V. N., & Stone, F. (1964). Changes in children's verbal behavior as a function of social approval, experimenter differences, and child personality. *Child Development 35(3) 1964, 669-676 Blackwell Publishing*.

Rozell, E. J., Gundersen, D. E., & Terpstra, D. E. (1998). Gender differences in the factors affecting helpless behavior and performance. *Journal of Social Behavior & Personality, 13*(2), 265-280.

Samuel, W. (1977). Observed IQ as a function of test atmosphere, tester expectation, and race of tester: A replication for female subjects. *Journal of Educational Psychology, 69*(5), 593-604.

Sarason, I. G., & Harmatz, M. G. (1965). Test anxiety and experimental conditions. *Journal of Personality and Social Psychology, 1*(5), 499-505.

Sclafani, J. D., & LaBarba, R. C. (1982). Sex differences and effects of sex of examiner in early conservation ability. *Bulletin of the Psychonomic Society, 19*(4), 191-193.

Scott, T., & Underwood, E. (1934). Sex of experimenter and intelligence as factors in maze learning: a further comparison of maze and sex difference. *Journal of Genetic Psychology 44 1934, 239-246 Heldref Publications*.

Senn, C. Y., & Desmarais, S. (2001). Are our recruitment practices for sex studies working across gender? The effect of topic and gender of recruiter on participation rates of university men and women. *Journal of Sex Research, 38*(2), 111-117.

Shilkret, C. J., & Masling, J. (1981). Oral dependence and dependent behavior. *Journal of Personality Assessment, 45*(2), 125-129.

Shope, G. L., Hedrick, T. E., & Geen, R. G. (1978). Physical/verbal aggression: Sex differences in style. *Journal of Personality, 46*(1), 23-42.

Shultz, T. R., & Hartup, W. W. (1967). Performance under Social Reinforcement as a Function of Masculinity-Femininity of Experimenter and Subject. *Journal of Personality and Social Psychology, 6*(3), 337-341.

Simon, W. E. (1971). Helping behavior in the absence of visual contact as a function of sex of person asking for help and sex of person being asked for help. *Psychological Reports, 28*(2), 609-610.

Singer, R. W., & Llewellyn, J. H. (1973). Effects of experimenter's gender on subject's performance. *Research Quarterly Vol 44(2) May 1973, 185-191*.

Smith, D. E., Gier, J. A., & Willis, F. N. (1982). Interpersonal touch and compliance with a marketing request. *Basic and Applied Social Psychology, 3*(1), 35-38.

Smith, J. L., Berry, N. J., & Whiteley, P. (1997). The effect of interviewer guise upon gender self-report responses as a function of interviewee's self-monitoring position. *European Journal of Social Psychology, 27*(2), 237-243.

Sobal, J., & Juhasz, J. B. (1977). Sex, experimenter, and reinforcement effects in verbal learning. *Journal of Social Psychology, 102*(2), 267-273.

Somervill, J. W., & et al. (1982). Differences in characteristics preferred by college students for academic advisors, vocational counselors, and psychotherapists: A preliminary report. *Perceptual and Motor Skills, 54*(1), 29-30.

Somervill, J. W., Mullenberg, S. A., Benz, B. L., & Chaisson, M. (1981). Sex of experimenter and effects of pretest interactions on approaching a snake by fearful female subjects. *Perceptual and Motor Skills, 53*(2), 403-406.

Somervill, J. W., Rimm, D. C., & Spudic, T. J. (1980). Sex of model and sex of subject in reduction of snake avoidance. *Perceptual and Motor Skills, 50*(1), 151-154.

Soukup, M. A., & Somervill, J. W. (1979). Effects of sex role and an audience on approaching a snake by mildly fearful subjects. *Perceptual and Motor Skills, 49*(2), 572.

Stark-Adamec, C., Adamec, R. E., & Pihl, R. O. (1982). Experimenter effects in marihuana research: A note of caution. *Psychological Reports, 51*(1), 203-206.

Stein, A. H., Pohly, S. R., & Mueller, E. (1971). The influence of masculine, feminine, and neutral tasks on children's achievement behavior, expectancies of success, and attainment values. *Child Development, . 42*(1), 195-207.

Stevenson, H. W. (1961). Social reinforcement with children as a function of CA, sex of E, and sex of S. *Journal of Abnormal & Social Psychology, 63*(1), 147-154.

Stevenson, H. W., & Odom, R. D. (1964). Visual reinforcement with children. *Journal of Experimental Child Psychology 1(3) 1964, 248-255 Elsevier Science*.

Stewart, D. J., & Resnick, J. H. (1970). Verbal conditioning of psychopaths as a function of experimenter-subject sex differences. *Journal of Abnormal Psychology, 75*(1), 90-92.

Tacon, P. H. (1965). The Effects of Sex of E on Obtaining Ss for Psychological Experiments. *Canadian Psychologist Psychologie Canadienne, 6a*(4), 349-352.

Takahashi, M. (2003). Recognition of odors and identification of sources. *American Journal of Psychology, 116*(4), 527-542.

Thomas, J. R., & Fortson, M. P. (1975). Experimenter effects on children's motor performance. *Journal of Motor Behavior, 7*(1), 65-72.

Thompson, J., & Connelly, J. J. (1988). Experimenter gender and size estimation accuracy. *International Journal of Eating Disorders, 7*(5), 723-725.

Van Hecke, M., & Tracy, R. J. (1983). Sex differences in children's responses to achievement and approval. *Child Study Journal, 13*(3), 165-173.

Voskuil, S., & Tucker, I. A. (1987). Examiner disability and other bias sources in administration of the Wechsler Adult Intelligence Scale--Revised. *Measurement & Evaluation in Counseling & Development, 20*(2), 89-96.

Walsh, R. T., & Schallow, J. R. (1977). The experimenter as a sex-role model in sex-role stereotyping. *Canadian Journal of Behavioural Science, 9*(4), 305-314.

Weisfeld, C. C., & Stack, M. A. (2002). When I look into your eyes: An ethological analysis of gender differences in married couples' non-verbal behaviors. *Psychology, Evolution & Gender, 4*(2), 125-147.

Weisse, C. S., Foster, K. K., & Fisher, E. A. (2005). The Influence of Experimenter Gender and Race on Pain Reporting: Does Racial or Gender Concordance Matter? *Pain Medicine, 6*(1), 80-87.

White, D. G. (1978). Effects of sex-typed labels and their source on the imitative performance of young children. *Child Development, 49*(4), 1266-1269.

Williams, S. W., Ogletree, S. M., Woodburn, W., & Raffeld, P. (1993). Gender roles, computer attitudes, and dyadic computer interaction performance in college students. *Sex Roles, 29*(7-8), 515-525.

Willis, F. N., & Hamm, H. K. (1980). The use of interpersonal touch in securing compliance. *Journal of Nonverbal Behavior, 5*(1), 49-55.

Woods, D. J., & Oppenheimer, K. C. (1980). Torque, hemispheric dominance, and psychosocial adjustment. *Journal of Abnormal Psychology, 89*(4), 567-572.