# Sexism, Statements, and Audits

Eric Schniter[a,b] and Timothy W. Shields[a,b, †]

[a]The Economic Science Institute, and
[b]Argyros School of Business & Economics
Chapman University

Version: 1 November 2014



**Abstract**

Gender stereotypes and gender-discriminant behaviors have been shown to have strong and undesirable organizational, managerial, and economic effects. We examine the relationship between sexism and accounting practices, and the effect of contextual feedback using laboratory experiments. Sexist stereotypes and contextual feedback may affect the likelihood of financial misstatements and audits when auditors and issuers are of known gender. To investigate these aspects of sexism at zero acquaintance and after contextual feedback, we presented males and females with incentivized belief elicitation tasks about anticipated interaction behaviors and then a series of strategic-communication game decisions in same, other, and unknown gender interactions. Feedback about belief accuracy, actual behaviors, and earnings was only given after completing a full set of belief elicitations and interactions. At zero acquaintance, both genders stereotyped the other gender's behavior propensities as relatively different than their own gender's. Both genders' stereotyped male and female targets similarly, and while both genders discriminated based on target gender, males' and females' behavior was similar. Consistent with a statistical discrimination account of sexism, stereotypes and game behaviors were adjusted after contextual feedback to more accurately reflect and predict others' behaviors. While biosocial and evolutionary perspectives may help explain why undesirable sexism is prevalent, our results suggest that by providing contextual information and incentives in reporting and auditing settings, we can motivate sexists to moderate their stereotypes and linked behaviors.

Keywords: sexism, gender, stereotype, discrimination, audit
JEL codes: A13, D03, J14, Z1

† Correspondence should be addressed to Timothy W. Shields, Argyros School of Business & Economics, Chapman University, One University Drive, Orange, California 92866; Telephone: (714) 289-2092; Email: shields@chapman.edu

## 1 Introduction

Gender organizes social relationships in all primate species (Smuts et al., 1987) and across human cultures (Sidanius, 1999). In the ancestral environments that shaped the design of human minds, knowing a person's gender would have supported inferences about their behavior in various selection-relevant contexts including those involving cooperative exchange, food production, mating, childrearing, conflict, and war (Buss, 1995; Geary, 1998). Across most human societies today, patriarchal dominance and the legacy of gender-specific divisions of labor, communication styles, and socialization norms further entrain our tendencies to think of the genders as distinct in their behavioral propensities and to behave in gender-appropriate ways (Bussey & Bandura, 1999). Whether based on ancient evolutionary or modern biosocial reasons, it is likely that another's gender is spontaneously encoded, such as when trading off conflicting motivations, predicting another's behavior, and deciding interaction behaviors.

If behavioral gender differences exist, but minimal information is available about a person with whom one needs to interact, stereotypes about that person may provide helpful guidance in choosing appropriate interaction behavior. Sexism, defined as behavior affect by gender stereotypes, could bring economic benefits by statistically discriminating on these; but often it does not. Sexism tends to be economically inefficient since it prevents the equalization of marginal rates of substitution in production that could otherwise be achieved (for a model of aggregate costs see Cavalvanti & Tavares, 2008**;** for estimates of actual costs see World Bank, 2007). Previous research has already uncovered various effects of gender and gender discrimination within accounting (Almer et al., 1998; Collins, 1993; Adams & Ferreira, 2009) and auditing professions (Gold et al., 2009, Hardies, Breesch & Branson, 2014; Francis et al., 2014.).[1] However, less is known about gender effects on reporting, and how gender stereotyping, gender-discriminate misstatement, gender-discriminate auditing, and contextual feedback about males' and females' behaviors interrelate.

We define stereotypes as "beliefs about the personal attributes of a social group" (Ashmore & Del Boca, 1981) and evaluate whether gender stereotypes about males and females are linked to gender-specific discrimination. Stereotypes of more communal and less agentic females (e.g.

[1] See also Hooks (1992), Anderson, Johnson, and Reckers (1994); Anderson-Gough, Grey, and Robson (2005), Kornberger, Carter, and Ross-Smith (2010) who have examined gender effects in the auditing workplace and career path.

Orbell, Dawes & Schwartz-Shea, 1994; Deaux & Lewis, 1984; Eagly, 2009)[2] scaffold common beliefs that they are more ethical (Mocan, 2008), kinder, more cooperative, and less willing to take risks than males (Balliet et al., 2011; Shelley et al., 2010; Eagly & Wood, 1999; Eagly & Crowley, 1986; Byrnes et al., 1999; Schubert, 2006; Eckel & Grossman, 2008b). Evidence that these beliefs are often held suggests that females should be expected less likely to deliberately issue financial misstatements and less likely to conduct financially risky audits, both behaviors that affect economic relationships. Archival data suggests that financial markets have responded to these gender stereotypes. Martin, Nishikawa, and Williams (2009) find market risk measures of companies following female CEO appointments to be significantly lower than after male CEO appointments, consistent with market participants believing that females (and the companies directed by them) are relatively risk averse and conscientious in their financial practices. Similarly, Krishnan and Parsons (2008) find that companies with more females in senior management are more profitable and have higher stock returns. Srinidhi, Gul and Tsui (2011) note that the less frequent misstatements and greater "earning quality" achieved when females' contribute to workplace diversity drives this profit effect. Ittonen, Miettinen and Vähämaa (2010) find that firms with a female audit committee chair have reduced auditing fees, and Niskanen et al. (2011) find that female auditors are more conservative in their reviews of reported income, both consistent with the belief that accounts of those firms under female control have a lesser inherent risk of financial misstatement.

Motivated by a rich gender differentiation literature in economic and behavioral sciences, we presented males and females with a set of incentive compatible belief elicitation tasks that we refer to as the "Guess game". Our Guess game has participants make predictions (that are later scored and paid for accuracy) about anticipated behavior in a strategic-communication game called the Bluff-Challenge game. In the Guess game participants predict the frequency of anticipated misstatements and audits in Bluff-Challenge game interactions with the "same" gender, "other" gender, and "unknown"-gender. These six guesses (2x3) inform us of personal and consensual stereotypes about the male and female genders at zero acquaintance.[3]

[2] A communal orientation conveys a concern for the welfare of others (e.g., Conway et al., 1996; Eagly & Steffen, 1984). In contrast, an agentic orientation involves a concern for one's own outcomes over others' outcomes (Campbell, Muncer & Gorman, 1993; Pruitt, 1983).

[3] Personal stereotypes are beliefs about groups (or beliefs about differences between groups) held by individuals. Consensual stereotypes are the aggregate beliefs about a particular group. Consensual stereotypes receive considerable attention among lay people and in the social science literature (e.g. Allport, 1979; Jost and Banaji,

Subsequent behaviors in the Bluff-Challenge game inform us of gender-based discrimination and whether discriminant behaviors are best responses to consensual stereotypes, personal stereotypes, or even economically profitable. Based on stereotype accuracy and corresponding discriminatory behavior we can evaluate whether participants practice "statistical" discrimination (Phelps, 1972; Arrow, 1973) by choosing best responses to accurate stereotypes or contextual feedback, or "taste-based" discrimination (Becker, 1957) that is independent of stereotype accuracy or others' actual behaviors. After receiving feedback about the accuracy of stereotypes elicited by the Guess game and results of interactions in the Bluff-Challenge game, we conducted a "Phase 2" in which the Guess game and Bluff-Challenge games were repeated to ascertain whether participants revise their stereotypes and/or gender discriminant behaviors in a "statistically" rational manner, based on feedback. These decisions and interactions were realized with college students in a laboratory using a computer-based economic experiment that provided us the control needed to investigate whether availability of information about another's gender might affect reporting and auditing practices.

In section 2 we review some background about the communication and auditing problems that our study models and then present our hypotheses. In section 3 we describe our method for studying sexism in the reporting and auditing context and in section 4 we present our results. Last, in section 5 we discuss results and conclude.

## 2 Background and Hypothesis

The act of making assertions is a fundamental communication problem. When issuing a financial statement, the issuer faces decisions about what to communicate and to whom. Compared to the wealth of knowledge on inter-gender communication in personal relationships (e.g., Gray, 1992) and in leadership relationships (Eagly & Johnson, 1990), less is known about how the delivery and receipt of financial information in strategic settings is affected by gender (but see Tannen, 1995). Signaling or communication games have been used by biologists, economists, and philosophers to study the transfer of information between individuals with conflicting interests. These researchers have sought to better understand how signals can evolve between senders and receivers of the same or different species (Searcy & Nowicki, 2005), how

1994; Pickering, 2001) because they represent the beliefs that are most widely shared and thought to be transmitted through discourse and popular media. Among stereotyped social categories (e.g., age, gender, ethnicity, and social class), gender is one of the most accurate consensual stereotypes (Jussim et al., 2009).

signal costs may be instrumental to signal honesty (Spence, 1973; Zahavi, 1975; Schniter, Sheremeta & Sznycer, 2013), and how conventional meaning arises in discourse (Lewis, 1969). In this paper we use a signaling game known as the "Bluff-Challenge game" (Schniter & Shields, 2014) that models a pair of fundamental communication issues affecting economic relationships: the decision whether to deliver a potentially profitable (or else more costly) misstatement about an economic state, and the decision whether to accept or challenge an unverified statement with a potentially costly (or else more profitable) audit.[4] While game theory provides a normative theory of behavior in this signaling game – useful as a comparative benchmark, expectations based on gender stereotypes affecting our study could produce best response behaviors (in terms of how misstatements are issued and challenged by audits) that deviate radically from Nash equilibrium behaviors.

We conceptualize an audit as a planned examination of statements and objective evidence about which statements were made for the purpose of evaluating if they are reasonably true and fair. Auditing is a crucial accounting mechanism that ascertains validity and reliability of a system's internal control over reporting quality. While many internal and external forms of auditing are intended as checks for mistakes, we focus on audits as veracity "challenges" intended to reveal deliberate misstatement and detect cheaters.[5] In particular, we abstract away from the notion of misstatements as accidents, for example due to computational or clerical errors, and instead examine a setting with economic incentive for deliberate misstatement. Just as auditors are important for economic relationships because they add credibility to financial reporting by expressing an opinion about the fairness of the financial statements, issuers of statements need to be trusted by investors so that stable and efficient economic relationships can prosper. In fact, companies with lower rates of financial misstatement and who employ auditors

[4] In our signaling game the sender (whom we refer to as the issuer) acquires private information about the state of the world and contingent on that information selects a signal (we call this a statement) to send to the receiver (whom we call the auditor). Having received the statement, and contingent on the possibility of misstatement, the auditor chooses whether to challenge the statement's veracity with additional auditing effort. As with most signaling games, payoffs for sender and receiver are functions of state of the world, action chosen, and the signal sent (Cho & Kreps, 1987).

[5] Evolutionary psychologists posit that humans have cognitive adaptations for cheater detection and hazard avoidance. Experimental evidence demonstrates that these adaptations direct attention to information enabling individuals to guess whether or not they are at risk (Cosmides & Tooby, 1992; Fiddick, Cosmides & Tooby, 2000). Neuropsychological evidence of selective impairment with these abilities demonstrates the association of cheater detection and precautionary reasoning with specialized brain areas (Stone et al, 2002).

charging higher fees (a proxy for audit effort) enjoy admired reputations of greater "earning quality" and greater market confidence (Cao, Myers & Omer, 2012).

## 2.1 Hypotheses

In considering possible gender effects on the strategic relationship between issuers of statements and auditors, we apply two lenses that provide us insight into possible reasons for gender differences in these behaviors: the biosocial and evolutionary perspectives. In considering possible effects of contextual feedback on gender stereotypes and discrimination, we apply another two explanatory frames: one "taste-based" and the other "statistical". Based on hypotheses advanced by these perspectives, we derive predictions of experimental behavior in zero-acquaintance settings where only gender is known, and after contextual feedback, where gender continues to be the only identifying attribute of interaction partner type.

### 2.1.1 Shared gender stereotypes at zero acquaintance.

Humans engage in categorization and construct stereotypes around constellations of traits believed to be true of individual members of a social category like gender (Ashmore & Del Boca, 1981). Stereotypes help people cope with social challenges, such as identifying, encoding, and recalling members of other groups, making sense of what qualifies a social group, and informing decisions of how to deal with them (e.g., Tajfel, 1981). One of these social challenges, relevant to our study, is managing cooperation at with members of a known gender in dyadic social dilemmas that provide incentive for non-cooperation.

"Mixed stereotypes" are sets of correlated stereotypes attributed to people based on their social category (Fiske, Cuddy, Glick & Xu, 2002). Mixed stereotypes about males and females have long been studied (e.g., Rosenkrantz et al., 1968; Broverman et al., 1972; Deaux & Lewis, 1984) and more recently shown across extensive cross-cultural samples (Glick et al., 2000; Lockenhoff et al., 2014). We consider a set of stereotypes about *cooperativeness* in our strategic-communication setting to involve a mix of "honest" and "trusting" attributes for cooperators, and a mix of "dishonest" and "challenging" attributes for non-cooperators.

The biosocial perspective hypothesizes that gender differences in behavior are most attributable to the social and cultural processes involved in their generation, and that a backdrop

of biological underpinnings sets the stage for social organization and culture.[6] Recurrent divisions of labor, reproduction and childrearing have led to standardized gender roles conserved and stereotyped across almost all foraging, horticultural, and agricultural societies (Wood & Eagly, 2002). As a consequence of these social and cultural processes, males are expected to develop more agentic behavioral traits (e.g., assertiveness, defiance, and control) that are suited for high status and leadership roles, while females develop more communal behavioral traits that are suited for domestic and interpersonal roles (Eagly & Crowley, 1986). Indeed, researchers find that the stereotypes of *communal* (cooperative, honest, and trusting) females and of *agentic* (competitive, deceptive, and challenging) males are shared across genders and cultures (Williams & Best, 1982). A biosocial perspective suggests social and cultural processes transmit these mixed stereotypes encountered among different groups of perceivers for simple and practical reasons; in so far as consensual stereotypes provide accurate descriptions of social groups, individuals lacking access to contextual experience and acquaintance with members of these groups (and therefore reliable personal stereotypes) can simply anchor their beliefs on these salient consensual benchmarks. The biosocially informed *gender role discrimination hypothesis* predicts that at zero acquaintance our participants will share these consensual stereotypes and conform to them.

The evolutionary perspective hypothesizes that gender differences in behavior are attributable to the evolved physiologies and psychologies that natural and sexual selection has conserved, and that any cultural effects on a gender's behavior must also ultimately stem from the evolved psychologies generating culture (e.g. Barkow, Cosmides & Tooby, 1992). While some adaptive problems in ancestral environments provided recurrent selection pressures in similar measure for males and females, the intensity of other adaptive problems—especially those determined by intersexual selection and intrasexual competition pressures—have differed between sexes.[7] As a consequence, natural and sexual selection have provided males and females

[6] Across human societies, gender roles have been constrained by biology: males and females tend to exhibit strength differences, size differences, and difference in terms of obligate reproductive costs (e.g., only females gestate and lactate). As a consequence of biological constraints and society's dominant modes of production, efficient divisions of labor have emerged. Evaluations of earlier forms of sociocultural or social constructionist (social role) theories of sex differences have suggest that a more balanced perspective should integrate these evolutionary theories (e.g. Archer, 1996), leading to this more modern social-biological synthesis.

[7] Buss (1995) identifies several sex-differentiated adaptive problems that humans have faced: paternity uncertainty, identifying reproductively valuable females, gaining sexual access to females, identifying males who are able to invest, identifying males who are willing to invest, coalitional warfare, and coalitional defense.

different psychological adaptations for dealing with their gender-specific problem sets (Geary, 1998). One relevant example of this concerns humans' natural history of foraging and exchange differences where men pursue high-variance large resources via hunting and where women forage for lower-variance smaller resources through gathering. These differences and complementarities in foraging strategies have been documented across known foragers by Marlowe (2007), appear to have been stable throughout evolutionary history, and continue to predict related gender differences in the modern environment (Krasnow et al., 2011)[8]. Indeed, various experimental studies find that males are more tolerant of risk and thus act more assertively when facing risky decisions.[9] The prediction we derive from this perspective is that males have a greater propensity to take economic risks by issuing deceptive misstatements and challenging statements with audits.

These two perspectives we consider, biosocial and evolutionary, lead us to our first set of predictions, A.1 and A.2, concerning gender-specific stereotypes about misstatement and audit propensity expected in the Bluff-Challenge game.

A.1. Both males and females will stereotype males as more likely than females to issue misstatements.

A.2. Both males and females will stereotype males as more likely than females to challenge statements with audits.

### 2.1.2 Gender interaction (issuer × auditor) effects on misstatement propensity.

The biosocially informed *gender role discrimination hypothesis* suggests that culture and society descriptively "assign" gender roles to males and females (e.g., in the form of *stereotypes*) that generate prescriptive expectations of gender-normative behavior and discrimination against deviants (Eagly, 2009). Deviation from these roles may come with social costs. For example, if females engage in counter-stereotypical behaviors, they may be evaluated negatively–especially

[8] The human evolutionary history of divided labor has led to evolved gender differences in foraging strategies where males and females preference tolerating different kinds of risks and exchange relationships (Kaplan et al., 2012), using different styles of navigation (Dabbs et al., 1998), and recalling and orienting towards spatially encoded details differently (New et al., 2007).

[9] Croson and Gneezy (2009, pg 450) review experimental evidence and report ten experiments where females are less likely to take on risks than males. Rau (2014) presents evidence that females are more loss averse resulting in a reluctance to sell capital losses. Similarly, Baldiga (2013) shows that females are less likely to guess at questions when there is a penalty for wrong answers. However, in settings with both risk and strategic interaction, the gender-specific evidence provides inconsistent support for a gender-differentiated account of female risk aversion and male risk tolerance (for reviews see Eckel & Grossman, 2008; Croson & Gneezy, 2009).

by their female peers (Costrich et al., 1975). This perspective hypothesizes that socially normalized roles for females' encourage females to develop more communal traits and males to demonstrate agentic traits.[10] As such, males are expected to be more competitive leading to prediction B.1.

B.1. Holding the gender of the auditor constant, males will be more likely than females to issue misstatements.

An evolutionary perspective hypothesizes that frequent *female intrasexual competition* over social and material resources for offspring (or high quality mates who can provide these) has selected for psychological and behavioral adaptations making females extraordinarily competitive with other females (Trivers, 1972).[11] As such, females are expected to behave more competitively and therefore less cooperatively with other females than with males.

An evolutionary perspective (e.g., Van Vugt's *male-warrior hypothesis*) applied to males suggests that frequent and violent intergroup conflict has selected for psychological and behavioral adaptations making males more competitive against out-groups and more cooperative within male coalitions.[12] As such, males are expected to behave more cooperatively with other males than with females, and male-male within-group interactions may be more prone to cooperation than female-female interactions (Balliet et al., 2011). Indeed, there is evidence for greater overall female cooperativeness,[13] extraordinary female-female competition,[14] and a pattern of male-male interactions exhibiting more cooperation than female-female interactions in experimental studies,[15] though a few exceptions are reported.[16]

---

[10] Nguyen and Ryan (2008), Fryer, Levitt and List (2008), and Charness and Rustichini (2011) provide evidence that participants conform to stereotypes when cued.

[11] Cross-cultural studies have demonstrated intensive (though not very violent) female aggression as a means of competition for mates and resources, and for defense (Burbank, 1987; Levinson, 1989).

[12] The *male warrior hypothesis* has been developed by Van Vugt and colleagues across several papers (Van Vugt, De Cremer & Janssen, 2007, Van Vugt & Park, 2009; and Van Vugt, 2011). Male groups are both the most likely perpetrators and victims of intergroup aggression, now and in the past (Van Vugt, 2011). This intergroup aggression has selected for specialized cognitive mechanisms that allow males to form alliances with other males. Anthropologists have provided rich descriptions of all-male groups that defend and pursue common interests, for example acquiring and protecting reproductive resources, and defending against the aggression of other groups (e.g., Otterbein, 1968; Otterbein & Otterbein, 1965; van Velzen & van Wetering, 1960; Chagnon, 1988).

[13] Evidence of overall greater female cooperativeness comes from studies by Bolton and Katok (1995), Eckel and Grossman (1998), Ellingsen, Johannesson, Mollerstrom and Munkhammer (2012), and Ortmann and Tichy (1999).

[14] Evidence female-female competition comes from Buss and Dedden (1990), Archer and Coyne (2005), Hess and Hagen (2006), and Vaillancourt and Sharma (2011).

[15] Evidence of greater male–male than female–female cooperation comes from Rapoport and Chammah (1965), Sutter, Bosman, Kocher, and Winden (2009), and Balliet, Macfarlan and Van Vugt (2011).

[16] A few exceptions fail to find predicted gender effects (e.g., Walters, Stuhlmacher & Meyer, 1998; Andreoni & Vesterlund, 2001; Eckel & Grossman, 2001; Dufwenberg & Gneezy, 2005).

Taken together, these evolutionary hypotheses suggest that males and females' behavior at zero-acquaintance is predicted to discriminate upon interaction partner's known gender such that both genders are expected more likely to issue misstatements to female auditors. Evidence from experimental studies supports this derived prediction, showing that both genders are more cooperative when they know they are interacting with males as opposed to with females.

Also, recall that the biosocially informed A.2 predicts both males and females to stereotype female auditors as less likely than male auditors to challenge issued statements. Given A.2 is supported, a *statistical discrimination hypothesis* expects a best response to trusting female auditors: to issue them more frequent misstatements. The above evolutionary hypotheses and the biosocially informed *statistical discrimination* hypothesis all lead us to our next set of predictions.

B.2. Compared to their interactions with female auditors, males are *less* likely to issue misstatements when interacting with male auditors.

B.3. Compared to their interactions with male auditors, females are *more* likely to issue misstatements when interacting with female auditors.

### 2.1.3 Gender interaction (issuer x auditor) effects on auditor challenge propensity.

The biosocially informed *gender role discrimination hypothesis* characterizes males as more competitive leading to prediction C.1.

C.1. Holding the gender of issuer interacted with constant, male auditors will be more likely than female auditors to challenge statements.

Also, recall that the biosocially informed A.1 predicts males are stereotyped as more likely than females to issue misstatements. Given A.1 is supported, a *statistical discrimination hypothesis* expects a best response to deceptive male issuers of statements: challenge them with audits more frequently. These lead us to prediction C.2.

C.2. All auditors will be less likely to challenge females' statements than to challenge males' statements.

The evolutionary perspective also brings us uniquely nuanced expectations.[17] The *male warrior hypothesis* expects that males have the propensity to behave more cooperatively with

[17] The evolutionary predictions C.3 and C.4 (that each gender challenges females more) are both counter to the biosocially informed *statistical discrimination* prediction C.1 (that each gender challenges males more).

other males than with other females, giving us prediction C.3. The *female intrasexual competition hypothesis* expects that females have the propensity to behave more competitively with other females than with other males, giving us prediction C.4.

C.3. Male auditors will be more likely to challenge females' statements than to challenge other males' statements.

C.4. Female auditors will be less likely to challenge males' statements than to challenge other females' statements.

### 2.1.4 Revision of gender stereotypes and discrimination after contextual feedback.

We investigate whether perceivers revise their gender stereotypes and discriminant propensities after being provided contextual feedback (information about previous stereotype accuracy and results of previous interactions with members of those genders).

Where stereotypes are already fairly *accurate* at zero acquaintance and where discriminant propensities are best responses to those zero acquaintance stereotypes (i.e. in equilibrium), we have less reason to expect noticeably different stereotypes or discriminant propensities after contextual feedback. On the other hand, where *inaccurate* stereotypes are shared among groups of perceivers at zero acquaintance, we can ascertain whether perceivers revise these stereotypes after being provided contextual feedback.[18] If inaccurate stereotypes are malleable and stereotypers exercise "statistical" stereotyping, we expect to see evidence of stereotype revisions. If stereotype revision toward contextual representativeness occurs and interaction behaviors remain unchanged, the revised stereotypes will more accurately predict interaction behaviors after feedback. This leads us to predictions D.1 and D.2. On the other hand, if people prefer not to revise their stereotypes and discriminate based on "taste", for example reflecting powerful consensual priors or fixed responses to distant adaptive problems, then we would expect the null of D.2.

D.1. Stereotypes after contextual feedback will differ from stereotypes at zero acquaintance.

D.2. Stereotypes after contextual feedback will be more representative of contextual feedback than stereotypes before feedback.

If gender discriminant propensities are malleable, statistical discriminators may choose to change these after receiving contextual feedback about previous interaction outcomes that were

[18] In experimental studies Andreoni and Petrie (2008) find stereotypes disappear when participants are provided information about the behavior others.

not maximally profitable. If both stereotypes and behavior are strongly informed by feedback, revised gender discriminant propensities should better accord with revised stereotypes and the contextual priors with which they are now acquainted. This leads us to predictions D.3 and D.4. If gender discriminant propensities at zero acquaintance were maximally profitable (and intended to be) or taste-based (i.e., made with a disregard for profit) we have reason to expect the null of D.3.

D.3. Gender discriminant propensities after feedback will differ from gender discriminant propensities at zero acquaintance.

D.4. After feedback, revision in gender discriminant propensities will correlate with revision in stereotypes.

Finally, though we do not expect participants to have awareness of or consciously pursue the 1/3 Nash equilibrium strategy for this game, we evaluate how closely they approach this benchmark. Indeed, Fundenberg and Levine (2009) have noted that as players progress through rounds of play and learn from experience, they often approach the Nash equilibrium as an indirect consequence of "trying to maximize their own payoff while simultaneously learning about the play of other agents". However, Waller and Felix (1984a, 1984b) examined auditor learning and found that experience was not sufficient, so it remains an open question whether participants demonstrate equilibrium behavior after contextual feedback in our study.

## 3 Method

Like other studies in accounting (e.g., Dopuch & King, 1911, 1992, 1996; Bloomfield, 1997; Zimbelman & Waller, 1999), the experiment places auditors and issuers in a strategic setting. The setting captures essential features of the auditing environment. Auditors are aware of the issuers' economic incentives to misstate their private information. Likewise, issuers are aware of the auditors' economic detriment for an incorrect acceptance of a misstatement and incorrect challenge of a true statement. If the auditor accepts a statement, the issuer receives a higher payoff when misstating. If the auditor challenges a statement, the issuer receives a higher payoff when not misstating.

In this section we describe participants, the computerized experimental tasks: (i) the Guessing game, and (ii) the Bluff-Challenge game, and the experimental procedure. To minimize experimenter demands or unintended social desirability effects on behavior by referring to roles

that might be associated with normative behaviors, we used the neutral terms 'sender' and 'receiver' to label issuer and auditor roles in our experimental tasks and materials.

### 3.1 Participants

80 students, randomly drawn from a subject pool consisting of graduates and undergraduates, participated in one of two sessions. We assured participants that they would not be deceived by experimenters and would have the opportunity to earn money based on their decisions and the decisions of others.

### 3.2 Guessing game

Our incentivized Guess game elicits stereotypes by asking participants to make guesses about how often target groups of senders (e.g. male or female), with whom they would interact, would send them deceptive misstatements and how often target groups of receivers would challenge their statement veracity. For control, participants made guesses about senders and receivers of "unknown" gender. These targets of unknown gender engage in interactions with members of "unknown" gender.

Guesses about both misstatement and challenge propensity were made on a scale ranging from never, 0% of the time, to always, 100% of the time. Participants made guesses about the anticipated interaction behaviors of senders and receivers from "same", "other", and "unknown"-genders participating in the experiment. The targets of these guesses would not interact with the perceivers directly, thereby precluding the possibility that guesses would have reason to directly influence the behaviors these perceivers would subsequently choose in Bluff-Challenge game interactions. To incentivize guesses we use a quadratic scoring rule,[19] where guessed and actual behaviors are within the unit interval. Participants are told they can earn money by guessing within 1/6th above or below the observed behavior. Our incentive compatible Guess game is easy to understand and the graphical and numeric feedback of guess accuracy provided an easy-to-interpret form of contextualized information about interactions between groups and their correspondence to stereotyped behaviors.

[19] Our quadratic scoring rule is maximum{0, $10(1 – 36(guess – actual)^2). This rule qualifies as a "proper scoring rule" (e.g., Aczel & Pfanzagl, 1967; Savage, 1971) or "strictly proper" scoring rule (Winkler & Murphy, 1968).

### 3.3 Bluff-Challenge game

The Bluff-Challenge game belongs to a class of strategic-communication games with asymmetric information called "Bluffing games" (Holm, 2010). In the game (see Figure 1), the issuer (S) witnesses a computerized coin-flip in which nature (N) first selects the state 2 or 4 with equal probability. The state determines the resources that are available to the issuer and auditor (R). The issuer observes the state and sends a message "two to split" or "four to split" to the auditor. If the state is 2, the true message "two to split" is sent, but if the state is 4, the issuer can either send the message "four to split" or else misstate by sending the false message "two to split". Upon seeing the message "two to split" the auditor must accept (A) or challenge (C) the message veracity. An auditor who sees the message "four to split" can only accept. When accepting a message, the auditor's payoff does not strictly depend upon the state, but it does upon the message: the auditor's payoff is 2 when the message is "four to split" and 1 when the message is "two to split". The auditor's payoff is 0 when challenging a true message, but is 4 when challenging a false message. The issuer's payoff is the state minus the auditor's payoff. By this design the issuer is economically incentivized to deceive the auditor by sending false messages and the auditor is economically incentivized to challenge deceptive messages.

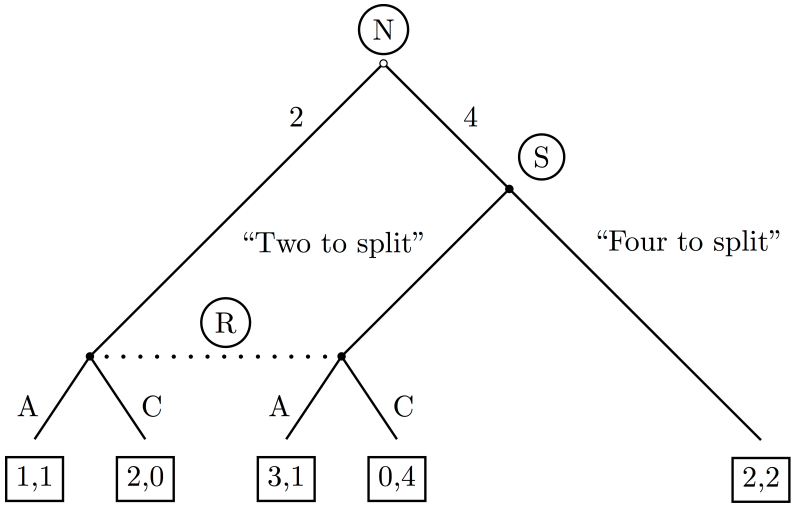



Figure 1. Bluff-Challenge Game Tree

While the Bluff-Challenge game bears some similarities to strategic-communication games of Gneezy (2005) and Sanchez-Pages and Vorsatz (2007), it differs by offering a single unique mixed equilibrium prediction that is interior (more than never but less than always): with our parameters, the issuer misstates a third of the time when the state is 4 and the auditor challenges a third of the time when the message is "two to split". We designed the game so that noise of random behavior (i.e., averaging to a half-of-the-time misstatement or rejection) would be discernable from a statistically apparent prediction.[20] These strategies result in expected payoffs of 5/3 for an issuer and 4/3 for an auditor. Unlike the aforementioned strategic-communication games, where the issuers always has misstatement opportunities, the opportunities for deceptive misstatement modeled in the Bluff-Challenge game arise occasionally by chance, are not affected by gender-dependent performance, and are equally available in "same", "other", or "unknown"-gender interactions – providing conditions for identifying when and how sexism endogenously occurs in auditing contexts. Likewise, while auditors in the Bluff-Challenge game do not always have opportunity to challenge, they can always elect a risk-free and cooperative option that gives a certain payoff.

### 3.4 Procedure

Participants, visually isolated from one another, are seated in individual workstations separated by partitions in a computer laboratory. Next, prerecorded audio and visual instructions are centrally presented to all participants in the laboratory (see Appendix B). After instructions, a quiz to ensure comprehension is administered and then the correct answers to the quiz are reviewed.[21] Questions are solicited and answered individually and privately before progressing with further tasks. All participants interact anonymously with each other via a web-based software interface using a computer network. Screenshots of the Guess game, feedback for the Guess game, and the Bluff-Challenge game are available in Appendix B. Participants are aware that there are two gender groups present and that they will be interacting with members of these groups throughout the session.

During Phase 1 participants make guesses (which we refer to throughout as "stereotypes") about the misstating and challenging propensities of targets from known or unknown genders.

[20] If parameter values for the states of nature (and therefore the payoffs) are changed, yet the economic incentives to deceive and to avoid deception remain, the game's mixed equilibrium changes—yet continues to remain interior.
[21] Quiz scores were not recorded and remained confidential.

Thereafter, participants interact in a series of dyadic strategic-communication interactions. All take the role of "Sender" (issuer) for eighteen interactions and thereafter take the role of "Receiver" (auditor) for eighteen interactions. The computerized software queues and distributes sender messages to appropriate receivers such that all participants take issuer and then auditor roles in the same order. Other than gender, personal identifiers are never revealed. Each participant is told whether their role and whether the person they interact with is of the (1) "same" gender, (2) "other" gender, or (3) "unknown" gender. With six interactions per group pairing per role (2×3×6), each participant interacted 36 times (with no feedback between interactions) per phase, and each time with a participant not previously interacted with in the same role. This allows us to evaluate whether participants are using mixed or pure misstatement and challenging strategies. At the conclusion of all 36 of these interactions, participants are informed about the computed accuracy of their stereotypes and results of their interactions, ending Phase 1. Phase 2 of the experiment repeats Phase 1, asking participants to make guesses about a different set of individuals than targeted by guesses in Phase 1, and to interact with participants whom they did not interact with under the same roles in Phase 1.

Each 40-participant session took approximately 80 minutes. A lottery was used to select a single guess and interaction from each phase for payment. Participants were paid at the conclusion of the experiment and earned an average of $19.63 (*SD* = $5.47) in addition to a fixed $7 payment for show-up.

## 4 Results

### 4.1 Overall results

We find no significant differences between sessions and report the combined results. At zero acquaintance, both males and females show sexist behavior with gender-specific stereotypes and gender-discriminant behavior. After contextual feedback, sexism is muted with uncooperative stereotypes attributed less frequently, and less evidence of gender discriminant behavior. These results are shown in Figure 2. In general, stereotypes about others of unknown gender (our control) lie midway between stereotypes about males or females, suggesting that an equal amount of stigmatism exists for male and female genders (though the characteristic attributions and effects of this may differ in many respects). We report results of stereotypes about and actual behaviors of the "unknown" gender group in Tables 1, 2, 3, and 4. For

simplicity and easier visual inspection we omit stereotypes about and behaviors of the "unknown" group from Figure 2.

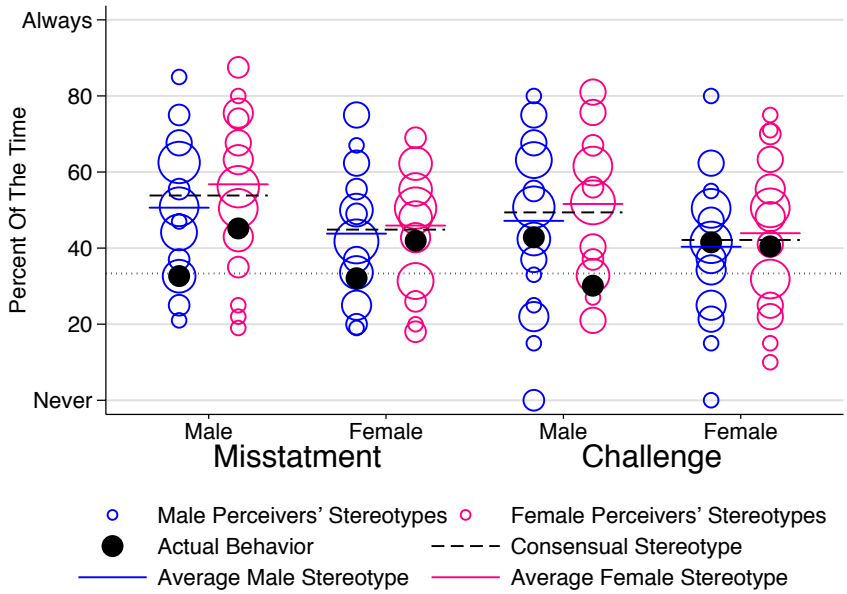
In Figure 2, we plot males' and females' stereotypes about targets, and the "actual" target behaviors by target gender. These actual behaviors reported are constructed by averaging all decisions by a participant in interactions with a target group (a) at zero acquaintance, and (b) after contextual feedback. All reported statistical tests consider a participant to be a single observation. Unless stated otherwise, reported results use Wilcoxon matched pairs test and significance is at the five percent level. Next, we present evidence for and against various hypotheses.

### 4.2 At zero acquaintance, do males and females share the same gender stereotypes of misstatement and audit propensity? Yes.

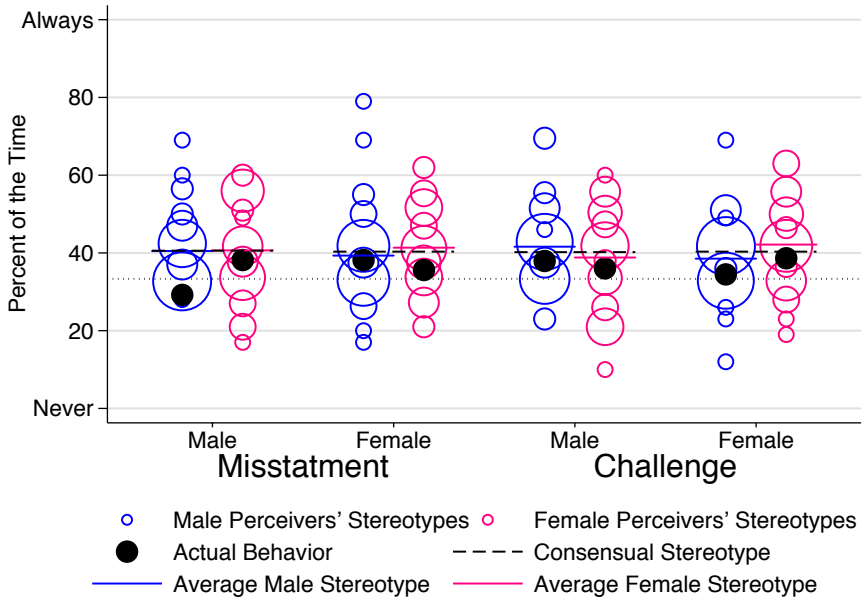
For each participant, we compare the guesses they made about the other gender's misstatement and audit propensities to the guesses they made about their same-gender's propensities. Table 1 reports these stereotypes at zero acquaintance and after contextual feedback. At zero acquaintance we find that males stereotype males as more likely than females to misstate ($Z = 2.37$, $p = .018$), and that females stereotype males as more likely than females to misstate ($Z = 3.91$, $p < .001$). We also find that males stereotype males as more likely than females to challenge at zero acquaintance ($Z = 2.55$, $p = .011$) and that females stereotype males as more likely than females to challenge ($Z = 3.21$, $p = .001$). Not only do both genders stereotype the opposite gender as relatively different, attributions of a target gender are shared among males and females who stereotype and discriminate based on target gender quite similarly: males are stereotyped as generally less cooperative (i.e., more likely to issue misstatements and challenging with audits) than females and treated accordingly.

We find no significant stereotype differences based on perceiver gender; stereotypes about male or female issuers or auditors are no different whether made by males or by females.

We find no significant stereotypes differenced based on the target gender after contextual feedback.




**(a) Bubble plot for *Phase 1* – at zero acquaintance**




**(b) Bubble plot for *Phase 2* - after contextual feedback**

**Figure 2: Bubble plots of stereotyped behavior propensities and actual behavior propensities at zero acquaintance and after contextual feedback.** The plot shows the distribution of perceivers' personal stereotypes about gender-specific issuer and auditor behavior, averages of perceiver's stereotypes, consensual stereotypes, and actual gender-specific issuer and auditor behaviors. The largest bubble represents 15 perceivers, the smallest bubble represents one perceiver. The dashed reference line at 1/3 of the time represents the mixed-strategy equilibrium prediction.

**Table 1. Perceiver's stereotypes at zero acquaintance and after contextual feedback**

| | Personal stereotypes about targets' propensities to | | | | | |
|---|---|---|---|---|---|---|
| | misstate given issuer's gender is: | | | challenge given auditor's gender is: | | |
| | Male | Unknown | Female | Male | Unknown | Female |
| *Perceiver at zero acquaintance* (*Phase 1*) | | | | | | |
| Male | 0.506 | 0.472 | 0.438 | 0.472 | 0.436 | 0.404 |
| | (0.024) | (0.023) | (0.024) | (0.032) | (0.025) | (0.024) |
| Female | 0.568 | 0.474 | 0.459 | 0.516 | 0.464 | 0.439 |
| | (0.026) | (0.023) | (0.022) | (0.027) | (0.025) | (0.026) |
| *Perceiver after contextual feedback* (*Phase 2*) | | | | | | |
| Male | 0.405 | 0.400 | 0.393 | 0.416 | 0.423 | 0.385 |
| | (0.015) | (0.012) | (0.018) | (0.016) | (0.015) | (0.015) |
| Female | 0.407 | 0.398 | 0.414 | 0.388 | 0.399 | 0.422 |
| | (0.019) | (0.014) | (0.017) | (0.020) | (0.015) | (0.017) |

*Note*: We report mean (standard error). All cells N = 40.

### 4.3 Are males more likely to issue misstatements? No.

For each issuer, we control for gender of auditor target and compare the actual propensity of males and females to issue deceptive misstatements. We do not have support for prediction B.1., since we fail to find any difference between male and female misstatement propensity based on the gender of auditor to whom statements were issued, whether at zero acquaintance, or after contextual feedback.

### 4.4 At zero acquaintance, are males and females more likely to issue misstatements to female auditors? Yes.

Next, for each issuer, we compare misstatement behavior when interacting with the same gender as opposed to the other gender. Table 2 reports misstatement behavior at zero acquaintance and after feedback. We find both genders tend to issue misstatements at zero acquaintance more often to female auditors than to male auditors. This difference is marginally significant for male issuers ($Z = 1.86$, $p = .062$) and significant for female issuers ($Z = 2.135$, $p = .033$). These results provide support for hypotheses B.2 and B.3. We fail to find any difference between male and female propensity to issue misstatements after contextual feedback, based on the gender of auditor.

**Table 2. Actual behavior propensities of males and females at zero acquaintance and after contextual feedback.**

| | Actual propensity to | | | | | |
|---|---|---|---|---|---|---|
| | misstate given auditor's gender is: | | | challenge given issuer's gender is: | | |
| | Male | Unknown | Female | Male | Unknown | Female |
| *At zero acquaintance* (*Phase 1*) | | | | | | |
| Male | 0.327 | 0.378 | 0.452 | 0.428 | 0.391 | 0.301 |
| | (0.045) | (0.045) | (0.049) | (0.046) | (0.043) | (0.041) |
| Female | 0.319 | 0.365 | 0.422 | 0.415 | 0.373 | 0.405 |
| | (0.047) | (0.040) | (0.054) | (0.048) | (0.039) | (0.050) |
| *After contextual feedback* (*Phase 2*) | | | | | | |
| Male | 0.291 | 0.328 | 0.382 | 0.379 | 0.306 | 0.360 |
| | (0.042) | (0.046) | (0.053) | (0.046) | (0.044) | (0.046) |
| Female | 0.385 | 0.338 | 0.360 | 0.345 | 0.380 | 0.386 |
| | (0.053) | (0.051) | (0.053) | (0.044) | (0.042) | (0.047) |

*Note*: We report mean (standard error). All cells N = 40.

### 4.5 Are male auditors more likely to challenge statements? No.

For auditors of each gender, we control for gender of issuer interacted with and compare the actual propensity of males and females to challenge with audits. We do not have support for the prediction that males are more likely to challenge statements (C.1), since we fail to find any difference between male and female propensity to challenge with audits based on the gender of issuer, whether at zero acquaintance, or after contextual feedback.

### 4.6 At zero acquaintance, which gender are auditors more likely to challenge? Neither.

Next, for each auditor, we compare challenge behavior when interacting with the same gender as opposed to the other gender. Table 2 reports auditor behavior at zero acquaintance and after contextual feedback. Male auditors tend to challenge females less than they challenge males, a marginally significant difference ($Z = 1.939$, $p = .053$). This result fails to support the evolutionary *male warrior* prediction (C.3), that males have the propensity to behave more cooperatively with other males than with other females (thus challenging other males less). For female auditors we fail to find a difference in the challenge of male issuers versus female issuers. As such, we fail to find support for the bioculturally informed *statistical discrimination* prediction (C.2) that all auditors will challenge male statements more often. We also fail to find support for the evolutionary *female intrasexual competition* prediction (C.4), that females have

the propensity to behave more competitively with other females (thus challenging them more).

### 4.7 Do misstatements and challenges correspond with personal stereotypes? Yes for males; No for females.

A more relaxed version of a *statistical discrimination hypothesis* predicts that best response misstatement and challenge behaviors will be made based on personal stereotypes about others' strategies (i.e, not necessarily based on biosocial or evolved consensual stereotypes). Indeed, the differences between male perceivers' stereotypes about male versus female auditors are negatively correlated to the actual misstatement differences that those male perceivers demonstrate when interacting with male versus female auditors (Spearman: $N = 40$, $\rho = -0.30$, $p = .060$). Likewise, the differences between male perceivers' stereotypes about male versus female issuers are positively correlated to the actual challenge differences that those male perceivers demonstrate when interacting with male versus female issuers ($N = 40$, $\rho = 0.47$, $p = .002$). Females' behaviors are not best-responses to the stereotypes they hold. Differences between female perceivers' stereotypes about male versus female auditors are not correlated to the actual misstatement differences that those female perceivers demonstrate when interacting with male versus female auditors (Spearman: $N = 37$, $\rho = 0.11$, $p = .520$). Likewise, differences between female perceivers' stereotypes about male issuers and female issuers are not correlated with different challenge behavior ($N = 40$, $\rho = 0.141$, $p = .385$).

### 4.8 Are gender stereotypes revised after contextual feedback? Yes.

Because stereotypes of gender propensities were inaccurate, and this inaccuracy often resulted in zero payment due to the scoring rule (average of 34 out of 80 guesses), we expect stereotypes to yield greater accuracy after contextual feedback. To evaluate whether perceivers revise their stereotypes about both the same gender's and other gender's behavior propensity based on contextual information about those groups' actual behavior, we compare individual stereotypes at zero acquaintance and after feedback (both reported in Table 1). After feedback, both males and females revise their stereotypes about each gender towards attributions of greater cooperative propensity. Males' ($Z = 3.617$, $p < .01$) and females' ($Z = 4.465$, $p < .01$) stereotypes about males' misstatement propensity are different after feedback. For stereotypes about males' challenge propensity, females ($Z = 3.483$, $p < .01$) show significant revision and males show

marginally significant revision (Z = 1.932, p = .053). Compared to the amount of revision seen with stereotypes about males, stereotypes about females undergo less revision. Stereotypes about females' misstatement propensity are marginally revised by males (Z = 1.844, p = .065) but not revised by females (Z = 1.533, p = .125). Stereotypes about females' challenge propensity are revised by neither males (Z = 1.341, p = .180) nor females (Z = 0.410, p = .681). As such we find partial support for the prediction D.1 that stereotypes are different after contextual feedback and this partial support only comes from the revision of stereotypes about males.

### 4.9 Are gender stereotypes more accurate after contextual feedback? Yes.

We evaluate the determinants of stereotype inaccuracy at zero acquaintance by conducting a two-way repeated measures analysis of variance that examines the personal stereotype discrepancies of perceivers according to perceiver gender and gender of target. The interaction term in this ANOVA captures the intergroup (same-gender other-gender) effect. We run this test twice for stereotypes at zero acquaintance, once for misstatement stereotype discrepancies and once for challenge stereotype discrepancies. For misstatement stereotype discrepancies, the main effect of target group is significant ($F(1,79) = 8.33$, $p < .01$, $\eta^2 = 0.048$) as is the main effect of perceiver group ($F(1, 79) = 8.64$, $p < .01$, $\eta^2 = 0.050$). However, the interaction (same-gender other-gender) effect is not significant ($F(1,79) = 0.09$, $p = .768$, $\eta^2 = 0.000$). These results stem from the fact that both males' and females' zero acquaintance stereotypes about males' misstatement propensity are significantly less accurate than their zero acquaintance stereotypes about females' misstatement propensity (see Table 3). For zero acquaintance challenge stereotype discrepancies, the main effects of target group ($F(1,79) = 18.47$, $p < .01$, $\eta^2 = 0.118$), and perceiver group ($F(1,79) = 15.82$, $p < .01$, $\eta^2 = 0.100$) are accurate, as is the interaction (same-gender other-gender) effect ($F(1,79) = 5.24$, $p = .023$, $\eta^2 = 0.033$). These results are due to (i) males' significantly less accurate stereotypes about males, relative to males' stereotypes about females, and (ii) females' less accurate stereotypes about both male and female challenge propensities, relative to males' stereotypes about male and female challenge propensities (see Table 3).

The most drastic changes in accuracy between Phases 1 and 2 are seen with stereotypes about male propensities to misstate and challenge. Table 3 also presents results of tests showing

that both males' and females' stereotype accuracy (for all target roles and genders) significantly improves after contextual feedback.

**Table 3. Stereotype accuracy at zero acquaintance and after contextual feedback**

| | Prediction error of stereotypes about | | | | | |
|---|---|---|---|---|---|---|
| | issuers whose gender is | | | auditors gender is | | |
| | Male | Unknown | Female | Male | Unknown | Female |
| *Perceiver at zero acquaintance (Phase 1) is* | | | | | | |
| Male | 0.195 | 0.140 | 0.151 | 0.167 | 0.133 | 0.112 |
| | (0.021) | (0.016) | (0.018) | (0.018) | (0.015) | (0.015) |
| Female | 0.165 | 0.143 | 0.124 | 0.230 | 0.149 | 0.144 |
| | (0.018) | (0.016) | (0.011) | (0.024) | (0.016) | (0.014) |
| *Perceiver after contextual feedback (Phase 2) is* | | | | | | |
| Male | 0.114** | 0.073** | 0.081** | 0.078** | 0.093* | 0.074* |
| | (0.015) | (0.011) | (0.013) | (0.012) | (0.013) | (0.011) |
| Female | 0.099** | 0.087** | 0.094* | 0.107** | 0.083** | 0.091** |
| | (0.012) | (0.011) | (0.012) | (0.010) | (0.012) | (0.011) |

*Note*: We report mean (standard error) of prediction error, defined as absolute value of prediction minus actual behavior. All cells N = 40. Accuracy significantly improved in Phase 2 (versus Phase 1): $p < .01$: ** $p < .05$: *

### 4.10 Do gender stereotype revisions appear to be influenced by contextual feedback? Yes.

To evaluate whether the above stereotype revision is influenced by contextual feedback (D.2), we first compare the revised stereotypes to contextual feedback. Table 4 reports the deviation between contextual feedback and Phase 2 stereotypes. The deviation measure after feedback is significantly different from zero for half of the stereotypes (see Table 4), suggesting that revision is not purely modeled on contextual feedback. That is, the revised stereotype is not simply a restatement of the contextual feedback, but appears to be based on a combination of contextual feedback and prior stereotypes. We compare the differences between zero-acquaintance stereotypes and contextual feedback (untabulated) to the differences summarized in Table 4. The differences between zero-acquaintance stereotypes and feedback are at least double the difference between post-feedback stereotypes and feedback. That is, revised stereotypes are closer to the contextualized feedback than are the stereotypes at zero-acquaintance.

There are plausible reasons to expect that stereotypes after contextual feedback are not simply equivalent to the feedback. First, stereotypes held at zero-acquaintance, while imperfect (as shown in Section 4.9), may not be completely invalid and without merit for reapplication in

Phase 2 when perceivers are making guesses about a new set of individuals that, though drawn from the same experimental context, they have not yet interacted with in those roles. The stereotypes about behavior made under Phase 1 "zero-acquaintance" conditions were likely informed by observations of numerous individuals encountered across the perceiver's lifetime of experiences, while the feedback given in this game is of a much smaller set of fewer individuals. Second, in Phase 2, participants are guessing about the behavior of a "moving target", who has also been given feedback. This moving target may have started the experiment with one set of behavior propensities, but upon receiving feedback about interactions may be inclined towards a new set of different propensities. Last, while participants where given contextual feedback, this feedback from Phase 1 was not visually available during Phase 2 when guesses and behavior decisions for interactions were again made (see screenshots in the appendix). As such, there is ample room for human error: even if individuals attempted to make post-feedback stereotypes exactly based on the contextual feedback viewed, they would have had to rely on their long term memory[22] and recall.

To ascertain how post-feedback stereotypes incorporated the contextual feedback, we examine the correlation between the difference between the contextual feedback and stereotype at zero-acquaintance (contextual feedback less Phase 1 guess), and the stereotype revision (Phase 2 guess less Phase 1 guess). The positive correlation is significant for all six stereotypes (about male, female or "unknown" issuers or auditors) held by both males and females (Spearman: $\rho \in [.615, .866]$, all $p < .01$). That is, all 12 revisions show a pattern of revising the stereotype downwards when the zero-acquaintance stereotype over-estimated the behavioral tendency, and upwards when the stereotype under-estimated the behavioral tendency. Stereotypes after feedback appear to be a combination of the original stereotype and the contextualized feedback. As such, we find support for the prediction D.2 that behavioral propensities stereotyped after contextual feedback incorporate contextual feedback, revising towards that feedback.

[22] While short-term memory span typically lasts not longer than 30 seconds for visually presented information (Posner & Konick, 1966), phase 2 stereotypes were elicited more than 30 seconds after the contextual feedback was last seen and no writing materials were made available (for participants to record feedback for later reference).

**Table 4. Deviation between contextual feedback about actual target propensities and stereotypes after contextual feedback.**

| | Deviation between stereotype about target and contextual feedback about: issuer's challenge propensity given their gender is: | | | auditor's misstatement propensity given their gender is: | | |
|---|---|---|---|---|---|---|
| | Male | Unknown | Female | Male | Unknown | Female |
| *Stereotype after contextual feedback* (*Phase 2*) | | | | | | |
| Male | -0.078** | -0.029* | -0.073** | 0.012 | -0.041* | 0.030† |
| | (0.015) | (0.012) | (0.018) | (0.016) | (0.016) | (0.016) |
| Female | 0.006** | -0.026 | 0.045 | -0.017** | -0.017 | -0.087 |
| | (0.021) | (0.014) | (0.015) | (0.017) | (0.017) | (0.019) |

*Note*: We report mean (standard error) of the value of stereotypes minus contextual feedback from Phase 1. All cells N=40. Significantly different from zero: $p < .01$: **, $p < .05$ : *, $p < .1$ : †

### 4.11 Does gender discriminant behavior change after contextual feedback? No.

We compare gender-discriminant behaviors at zero acquaintance and after feedback (both reported in Table 2). We find no significant differences in averaged behavior between phases, and no support for the prediction that discriminant propensities change after contextual feedback (D.3).

However, we do find evidence that behavior after feedback approaches the predicted equilibrium where players misstate and challenge ⅓ of the time. Before feedback (Table 2 Phase 1) we can reject that both males and females misstate a ⅓ of the time when interacting with female auditors and misstate ½ of the time when interacting with male auditors. Likewise we can reject that before feedback both males and females challenge ⅓ of the time when interacting with male issuers, and challenge ½ of the time when interacting with female issuers. We can reject that behavior after feedback (Table 2 Phase 2) is simply random (i.e., averaging to ½ of the time). At the same time, we fail to reject that participants misstate and challenge ⅓ of the time after feedback —the mixed strategy equilibrium prediction.

### 4.12 Does gender discriminant behavior after contextual feedback correlate with stereotype revisions? Yes.

We evaluate whether individuals' changes in gender discriminant propensities (from Phase 1 to Phase 2) are correlated with stereotype revisions. Below we report the correlation between revisions in misstatement (challenge) behavior and revisions in challenge (misstatement) stereotypes. When the target interacted with was male, we find the following significant results

using Spearman's correlation test. Participants' revision of misstatement behavior and revised challenging stereotypes is negatively correlated (N = 75, $\rho = -.384$, p < .01), consistent with a best response: misstating more to male auditors as a consequence of a stereotype revised towards attribution of their less frequent challenge propensity.[23] The revision of challenging behavior and revised misstatement stereotype is positively correlated (N = 80, $\rho = .279$, p = .012), consistent with a best response by both genders: challenging the target less as a consequence of a stereotyped revised towards attribution of a lesser misstatement propensity. When the target was female, we do not find significant correlation between revised misstatement behavior and revised challenge stereotype (N = 77, $\rho = -.042$, p = .716), nor do we find a significant correlation between revised challenging behavior and revised misstatement stereotype (N = 80, $\rho = .050$, p = .658). Thus, we only find support for D.4 when the target is male.

As reviewed above, post-feedback stereotypes tend towards contextual priors and, compared to zero-acquaintance discrimination, discriminant behaviors show little difference after feedback across the board. However, when there is a behavioral revision, post-feedback discriminant behavior tends towards a best response to revised stereotypes.

## 5 Discussion and Conclusion

While there has been extensive investigation of male and female economic behaviors in the laboratory, evidence of behavioral differences is mixed and inconsistent. Specifically missing from the literature on gender effects is experimental evidence of a link between gender differences and gender stereotypes, and evidence of a link between gender stereotypes and gender discrimination. In this paper we reported results demonstrating sexism in the laboratory and evaluating the accuracy and economic consequences of gender-discriminant beliefs and behaviors. Below we discuss implications of what we have learned about the links between genders, sexist beliefs, sexist behaviors, and effects of contextual feedback.

At zero acquaintance, there is a positive correlation between stereotypes about a gender's propensity to issue a misstatement and that gender's propensity to challenge a statement with an audit – indicating that these stereotypes more broadly reflected a latent dimension of (un)cooperativeness attributed to genders. Both genders stereotyped males as relatively

[23] In order to measure a revision of misstatement behavior, a participant had to see at least one coin flip of `4' in each phase. This excludes five subjects when interacting as issuers with male auditors, and three subjects when interacting as issuers with female auditors.

uncooperative (8.8% of the time more likely to attempt deception and 7.2% of the time more likely to challenge message veracity) and stereotyped females as relatively cooperative. Furthermore, both male and female participants stereotyped the uncooperativeness of the "other" group (the other gender) as relatively different than their own (same-gender) group (an average difference of 8%). While there was consensus regarding stereotypes elicited at zero acquaintance according to target-gender, and the average (i.e. "consensual") stereotype produced by the aggregate tended to demonstrate "wisdom of the crowd" by having greater accuracy than most participants' personal stereotypes, our experiment reveals that those consensual stereotypes were relatively inaccurate and inferior when compared to contextually informed consensual stereotypes.[24] We fail to find a gender difference in behaviors and also fail to find a same-gender-other-gender basis for behavior differences across interactions. Both genders show a tendency to inaccurately predict males' behaviors. Holding constant the target's gender, males and females do not differ in how often they choose to issue misstatements and challenge statements with audits, however both genders do show sexist discrimination issuing more misstatements to female auditors than to male auditors.

Participants adjust their stereotypes and game behaviors after feedback to more accurately reflect contextual priors. Because discriminant behaviors do not change drastically after contextual feedback, revised stereotypes predict others' behaviors better than zero acquaintance stereotypes. After feedback both genders show less sexism in their gender stereotyping with males and females stereotyped more similarly and significantly more accurately than at zero acquaintance. Additionally, after feedback, neither gender nor gender interaction affect behavior: across interactions, the genders do not differ in their propensity to issue misstatements and challenge statements with audits. In fact, the behavior observed after feedback quite closely approaches the Nash equilibrium prediction, suggesting that participants are utilizing learning mechanisms that integrate new contextual information for the purpose of updating prior gender stereotypes to more accurately reflect contextual attributions.

[24]"Wisdom of the crowd" was demonstrated where consensual stereotypes about gender groups corresponded to actual behaviors better than did most individuals' personal stereotypes: 69% of personal stereotypes about males' uncooperativeness (specifically 58% of issuer misstatement stereotypes, 81% of auditor challenge stereotypes, N = 80) were inferior and 84% of personal stereotypes about females' uncooperativeness (specifically 71% of issuer misstatement stereotypes, 96% of auditor challenge stereotypes, N = 80) were inferior. "Wisdom of crowds" was has also been reported for ageist stereotypes produced with similar experimental tasks (Schniter & Shields, 2014).

Most experimental laboratory studies (e.g., Charness, Rigotti & Rustichini 2007; Sutter, 2009; Hargreaves-Heap & Zizzo, 2009; Chen & Li, 2009; Chen & Chen, 2011) examining intergroup bias have focused on minimal group designs, employing arbitrary "labeling" to induce group membership and behavioral effects (e.g., increased cooperation with in-group members). Compared to minimal group studies, the few experimental studies conducted on real social groups (Goette, Huffman & Meier, 2012; Durlauf, 1999; Bernhard, Fehr & Fischbacher 2006; Schniter & Shields, 2014) have yielded comparatively different results. Despite the advantages of convenience and experimental control, minimum group designs may be poor models of the real social groups that people respond to in unique ways. Inconsistent with intergroup bias results from minimal group designs, we find that males stereotype members of their in-group as least cooperative (Table 2). At zero acquaintance, male perceivers are marginally more accurate when stereotyping female versus male issuers ($Z = 1.741$, $p = .082$). Overall, female perceivers are significantly more accurate ($Z = 2.117$, $p = .034$). Holding the gender of the issuer fixed, Wilcoxon rank-sum tests show no significant differences in stereotype accuracy based on perceiver gender. At zero acquaintance, both male and female perceivers are significantly more accurate when stereotyping female versus male auditors (male: $Z = 2.277$, $p = .026$; female: $Z = 2.777$, $p < .01$). Holding the gender of the auditor fixed, male perceivers are more accurate than female perceivers when stereotyping male auditors (Wilcoxon: 1.946, $p = .049$) and female auditors (Wilcoxon: $Z = 2.128$, $p = .033$). At zero acquaintance both males and females act relatively uncooperative with female auditors, issuing them misstatements more often (see Table 2).

We find mixed evidence that stereotypes strategically inform behaviors and that stereotype inaccuracy leads to unprofitable discriminant behavior. Compared to the frequency of honest statements they issue their own gender, males are relatively less cooperative, issuing more misstatements when interacting with females they incorrectly believed to be relatively cooperative. Likewise, consistent with the belief that females are more cooperative, males challenged females' statements with audits less than they did males' statements. For males, this combination of beliefs about females and female discriminant behavior is directionally consistent with a best response intent on exploitation, deception avoidance, and profit maximization. On the other hand, female auditors were not relatively less cooperative with male issuers of statements, despite stereotyping males as more likely than females to issue misstatements. As such, females'

behavior is not consistent with the belief-informed money-maximizing best response expected of "statistical discrimination" (Arrow, 1973), but instead consistent with a form of costly "taste based discrimination" (Becker, 1971).

Our research contributes to the documentation of gender-effects in financial reporting and auditing contexts.[25] We contribute to this literature by directly assessing participants' stereotypes about genders they interact with, and directly measuring behavior conditionally upon the gender participants are interacting with, and thus providing a direct measure of discrimination. Studies examining archival data cannot measure rates of deception and successful deception, nor assess how often auditors challenge and the consequences of this non-observable challenge. Beyond not being able to provide direct measures of behavior, choices by issuer and auditors over whom to employ, and whom to serve, entail issues of endogeneity. Experimental studies of individual decision-making can measure gender differences in behavior, but cannot examine interaction between the genders. In our incentivized study, we can assess whether behavior is strategic conditional upon the stereotypes held, whether those stereotypes are accurate, and if stereotypes are malleable.

We also contribute to the accounting literature by introducing into consideration the complementarity of established biosocial and evolutionary perspectives for predicting gender differences in misstatement and auditing behavior.[26]. Rather than relying on documented gender differences in unrelated environment as a basis for predicting differences in our experimental environment, we generated our predictions from the principles of established evolutionary and biosocial theories that explain why gender differences are conserved, why gender stereotypes may be formed, and in what settings they are likely to manifest.

Auditors add credibility to financial reporting by expressing an opinion about the fairness of the financial statements. Likewise issuers of statements (e.g. about the success or valuation of stocks or a company's various forms of productivity and investment) need to be trusted by investors. Other researchers have claimed that a change in an institution's gender composition can lead to better auditing and reporting due to persistent gender stereotypes, e.g. that female

[25] There are two streams to this literature: one based on archival data (e.g. see Ittonen, Vahamaa & Vähämaa, 2013; Thiruvadi & Huagn, 2011; Sun, Liu & Lan, 2011; Barua, Davidson, Rama & Thiruvadi, 2010; Gul, Srindhi, & Tsui, 2011; Abbot, Parker & Presely, 2012; and Gul, Hutchinson & Lai, 2013), the other on experimental data (e.g. see Chung & Monroe, 1998; O'Donnell & Johnson, 2001; and Gold, Hunton & Gomaa, 2009).

[26] The complementarity of biosocial and evolutionary perspectives has already been suggested in discussions outside of accounting (Kenrick, 1987; Campbell, 1999; Eagly & Wood, 1999).

auditors are more diligent in uncovering misstatement, and that females are less likely to issue misstatements. However, we do not find gender differences in behavior, so are skeptical of the claims that woman directors are better monitors due to their ethical values and greater risk aversion (Gul, Hutchinson & Lai 2013). Other studies have suggested higher audit fees charged by female auditor engagement partners may be indicative of higher audit effort due to risk tolerance and/or higher diligence (Ittonen & Penn, 2012). Our findings suggest that females tend be the targets of deception attempts more than males, and that both males and females attempt to deceive females more by issuing misstatements. Our findings suggest that the differences noted in audit investment by female engagement partners may actually be the result of differential issuer behavior. That is, professional females may have learned to challenge issuers' assertions because they face higher rates of attempted deception. In short, while we would like to see a more fair integration of women into male dominated professions like accounting, we are unsure whether this would bring about changes in behavior.[27]

One possibility is that experience interacting with the other gender would bring about more homogeneous beliefs and behaviors for males and females. Game theory predicts that rational agents in our Bluff-Challenge game will use mixed strategies in equilibrium. The behavior we observed at zero acquaintance does not match the equilibrium prediction. This should not come as a surprise, though, as others have shown that mixed strategy predictions have low predictive power (e.g., Bloomfield, 1994; Erev & Roth, 1998). The behavior we observed after contextual feedback was quite different: it mirrored the equilibrium prediction of ⅓. Furthermore, we can reject that this behavior was simply random (i.e, with an expected average of ½).

We consider our findings valuable when viewed in light of other laboratory and field experiments that have considered the consequences of gender on auditor-issuer interactions. However, there are limitations in extrapolating results beyond our experiment. First, the participants were randomly selected from a pool consisting of primarily undergraduate students. While our experimental tasks do not require special knowledge or expertise, there is always the possibility that professionals who have self selected into auditing and management may not have

[27] The American Institute of CPAs (AICPA) Women's Initiatives Executive Committee report that while woman represent roughly half of the newly certified CPAs, woman represent only one-fifth of the partners in CPA firms and roughly one-third of AICPA general membership. Woman represent one-seventh of board directors and less than one-tenth of all CFOs (AICPA, 2014).

the same distributions of stereotypes, discriminatory behavior, and risk tolerances as our pool. Future studies may want to duplicate ours using a sample of professionals.

Second, in our study there was a single individual acting as issuer and as auditor. While in our experimental task the gender of the individual issuer or auditor was salient, we do not expect that professionals conceptualize auditing firms, or their issuer clients, as individuals with discrete gender identities. Rather than modeling institutional relationships, our strategic interaction more closely models the one-to-one interactions that can occur when an accounting staff person audits an individual's expense reports, or when an Internal Revenue Service auditor examines the return of an individual taxpayer.

Last, we looked to theories of why one might expect to see gender differences in behavior or gender-discriminant behavior in an auditor-issuer setting and found support for these theories when examining issuer behavior. However, when examining auditor behavior, we did not find support for the *biosocially informed discrimination hypothesis*, the *male warrior hypothesis*, or the *female intersexual competition hypothesis*. The lack of support for these hypotheses may be attributed to the absence of important cues necessary to trigger the predicted auditor response. For example, the prescription of normative behavior expected according to the *biosocially informed discrimination hypothesis* may hinge on cues informative of others' expectations and the threat of consequences for not meeting those expectations. Males' preference for cooperation among males predicted by the *male warrior hypothesis* may be triggered only in the presence of cues that one is part of an all-male coalition, facing threats from rival coalitions. Females relatively uncooperative treatment of other females predicted by the *female intersexual competition* may be triggered only in the presence of cues that limited resources are being competed over by rival females in a situation that cannot be improved upon by cooperation. The experimental task did not cue prescriptive behavior by disclosing others' stereotypes, did not involve coalitionary competition or threats, and did not present contest over limited resources. Future work that makes these cues salient by altering the setting may be more helpful for ascertaining where these hypotheses are better applied.

Future research may also wish to include field studies of professional auditors and their clients. For example, greater confidence could be placed in our findings by giving our experimental tasks to a convenience sample of professionals and finding external validity for the main results. Likewise, survey data of professional auditors might inquire as to the magnitude

and frequency of adjustments (waived or booked) while controlling for the gender of the engagement manager and the gender composition of the issuer firm. Finally, longitudinal studies can examine if the institution of a training program designed to reveal auditor sexism and provide an accuracy measure, result in more efficient auditing (e.g., as measured by restatement propensity, budget over- and under-runs, earnings quality, etc.).

Our work compliments the research on implicit attitudes (attitudes people hold without consciously being aware of them) that have been shown to predict discriminate behavior (e.g., see Bazerman, Morgan, and Loewenstein, 1997 and Bazerman, Loewenstein, and Moore, 2002). Using computers, experimenters have applied the Implicit Association Task paradigm to study implicit attitudes. Similarly, our computer-based task could serve as a diagnostic measure of gender discrimination in the auditing domain, decreasing audit firm's vulnerability to this unconscious bias.

Our results suggest that while the problem of harmful sexism may be common in an accounting and auditing context, it should be thought of as avoidable. Recognizing the many possible psychological and social reasons to expect sexism, we wish to call attention to the message that sexists can be motivated to change their beliefs, so as to bring improvements to themselves and those they affect. Research on racism and sexism argues that these undesirable traits can be remedied with decategorizing frames and counter-stereotype training processes, suggesting that similar results might be achieved for sexism. "Decategorization" processes (Brewer & Miller, 1984) seek to eliminate categorization by not allowing differentiation based on social category. For example, Kurzban, Tooby, and Cosmides (2001) have demonstrated that the automatic (implicit) encoding of race after brief presentations is attenuated when racial identity is orthogonal to meaningful group membership. They achieved this by using new minimal group frames to identify alliances. These "arbitrary" frames essentially trumped the role of race (a "real" social classification which people have a lifetime of experience using) in forming social group identity. Additionally, "counter-stereotype training" (Kawakami, Dovidio & van Kamp, 2005, 2011) and intervention programs designed to reduce gender-bias in hiring practices (Isaac, Lee & Carnes, 2009) have been shown to be successful in correcting undesirable gender stereotypes and decreasing intergroup bias in hiring decisions. Reuben, et al. (2013) find that perceivers update their gender stereotypes and reduce discrimination after being provided full information about the target gender's previous performance. Likewise, we found that under

incentivized conditions, participants used contextual information to adjust their stereotypes. We suspect similar approaches countering undesirable sexist stereotypes holds much promise and can be pursued by auditing firms in an effort to improve risk assessment.

## 6 References

Abbott, L. J., Parker, S., & Presley, T. J. (2012). Female Board Presence and the Likelihood of Financial Restatement. *Accounting Horizons*, *26*(4), 607–629.

Abdolmohammadi, M., & Wright, A. M. (1987). An Examination of the Effects of Experience and Task Complexity on Audit Judgments. *The Accounting Review*, *62*(1), 1–13.

Aczel, J., & Pfanzagl, J. (1967). Remarks on the measurement of subjective probability and information. *Metrika*, *11*(1), 91–105–105.

Adams, R., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94 (2): 291–309.

Allport, G.W., (1979). *The Nature of Human Prejudice*. Basic books.

AICPA, (2014), Womans Initiatives Executive Committee Brochure. Retrieved from http://www.aicpa.org/career/womenintheprofession/downloadabledocuments/10958-378_wiec%20brochure_final.pdf

Anderson, J. C., Johnson, E. N., & Reckers, P. M. J. (1994). Perceived effects of gender, family structure, and physical appearance on career progression in public accounting: A research note. *Accounting, Organizations and Society*, *19*(6), 483–491.

Anderson-Gough, F., Grey, C., & Robson, K. (2005). "Helping them to forget.": the organizational embedding of gender relations in public audit firms. *Accounting, Organizations and Society*, *30*(5), 469–490.

Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology*, 29(1), 73–93.

Andreoni, J., & Vesterlund, L. (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics*, *116*(1), 293–312.

Archer, J. (1996). Sex differences in social behavior: are the social role and evolutionary explanations compatible? *American Psychologist, 5*1(9), 909.

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review, 9*(3), 212-230.

Arrow, K. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees, *Discrimination in Labor Markets* (Vol. 3, pp. 3–33). Princeton, New Jersey: Princeton University Press.

Ashmore, R. D. & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D.L. Hamilton, *Cognitive Processes in Stereotyping and Intergroup Behavior* (pp. 1-35) Hillsdale, NJ: Erlbaum.

Baldiga, K. (2014). Gender Differences in Willingness to Guess. *Management Science*, 60(2), 434–448.

Balliet, D., Li, N.P., Macfarlan, S.J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*(6), 881–909.

Barkow, J. H., Cosmides, L. E., & Tooby, J. E. (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* Oxford University Press.

Barua, A., Davidson, L. F., Rama, D. V., & Thiruvadi, S. (2010). CFO Gender and Accruals Quality. *Accounting Horizons*, *24*(1), 25–39.

Bazerman, M. H., Loewenstein, G., & Moore, D. A. (2002). Why good accountants do bad audits. *Harvard business review*, 80(11), 96-103.

Bazerman, M. H., Morgan, K. P., & Loewenstein, G. F. (1997). The impossibility of auditor independence. *Sloan Management Review*, 88.

Becker, G. S. (1957). *The economics of discrimination*. Chicago: University of Chicago Press.

Ben-Ner, A., Kong, F., & Putterman, L. (2004a). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, *25*(5), 581–589.

Ben-Ner, A., Putterman, L., Kong, F., & Magan, D. (2004b). Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization*, *53*(3), 333–352.

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *The American Economic Review*, 217-221.

Bloomfield, R. (1994). Learning a mixed strategy equilibrium in the laboratory. *Journal of Economic Behavior & Organization*, *25*(3), 411–436.

Bloomfield, R. (1997). Strategic Dependence and the Assessment of Fraud Risk: A Laboratory Study. *The Accounting Review*, 72(4), 517–538.

Bolton, G. E., & Katok, E. (1995). An experimental test for gender differences in beneficent behavior. *Economics Letters*, 48, 287–292.

Brewer, M. B., & Miller, N. (Eds.). (1984). Groups in contact: The psychology of desegregation. Academic Press.

Broadbent, J., & Kirkham, L. (2008). Glass ceilings, glass cliffs or new worlds?: Revisiting gender and accounting. *Accounting, Auditing & Accountability Journal*, *21*(4), 465–473.

Broverman, I. K., Vogel, S. R. Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues, 28*(2), 59-78.

Burbank, V. K. (1987) Female aggression in cross-cultural perspective. *Behavior Science Research 21*:70–100.

Buss, D. M. (1995). Psychological sex differences: Origins through sexual selection. *American Psychologist*, 50, 164-168.

Buss, D. M., & Dedden, L. A. (1990). Derogation of competitors. *Journal of Social and Personal Relationships, 7*(3), 395-422.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological review*, 106(4), 676.

Cavalcanti, T. V. , & Tavares, J. (2008). The Output Cost of Gender Discrimination: A Model-Based Macroeconomic Estimate, Proceedings of the German Development Economics Conference, Zürich 2008, No. 43

Campbell, A. (1999). Staying alive: Evolution, culture, and women's intrasexual aggression. *Behavioral and Brain Sciences*, *22*(02), 203-214.

Campbell, A., Muncer, S., & Gorman, B. (1993). Sex and social representations of aggression: A communal-agentic analysis. *Aggressive Behavior*, *19*(2), 125–135.

Cao, Y., Myers, L. A., & Omer, T. C. (2012). Does Company Reputation Matter for Financial Reporting Quality? Evidence from Restatements. *Contemporary Accounting Research*, 29(3), 956-990.

Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science, 239*(4843), 985-992.

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual Behavior and Group Membership. *The American Economic Review*, *97*(4), 1340–1352.

Charness, G., & Rustichini, A. (2011). Gender differences in cooperation with group membership. *Games and Economic Behavior*, 72(1), 77–85.

Chen, R., & Chen, Y. (2011). The potential of social identity for equilibrium selection. *The American Economic Review*, *101*(6), 2562-2589.

Chen, Y. & Li, S. X. (2009). Group Identity and Social Preferences. *American Economic Review*, *99*(1), 431–57.

Chin, C.-L., & Chi, H.-Y. (2008). Gender Differences in Audit Quality. *2008 Working Paper.*

Cho, I-K. & Kreps, D. M. (1987) Signaling games and stable equilibria. *Quarterly Journal of Economics,*102:179-221

Chung, J. (2002). Performance evaluation relating to mas decisions: rater age, ratee sex and level of performance. *Advances in Public Interest Accounting*, *9*, 45–63.

Chung, J., & Monroe, G. (1998). Gender differences in information processing: An empirical test of the hypothesis-confirming strategy in an audit context. *Accounting and Finance*, *38*(2), 265–279.

Chung, J., & Monroe, G. S. (2001). A Research Note on the Effects of Gender and Task Complexity on an Audit Judgment. *Behavioral Research in Accounting*, *13*(1), 111–125.

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25–38.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In: J. Barkow, L. Cosmides, and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford Unviersity Press: New York, pp.163-228

Costrich, N., Feinstein, J., Kidder, L., Marecek, J., & Pascale, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology*, 11(6), 520–

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Czarniawska, B. (2008). Accounting and gender across times and places: An excursion into fiction. *Accounting, Organizations and Society*. *33*(1), 33–47

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*(5), 991–1004.

Dopuch, N., & King, R. R. (1991). The Impact of MAS on Auditors' Independence: An Experimental Markets Study. *Journal of Accounting Research*, 29, 60–98.

Dopuch, N., & King, R. R. (1992). Negligence versus Strict Liability Regimes in Auditing: An Experimental Investigation. *The Accounting Review*, 67(1), 97–120.

Dopuch, N., & King, R. R. (1996). The Effects of Lowballing on Audit Quality: An Experimental Markets Study. *Journal of Accounting, Auditing & Finance*, 11(1), 45–68

Dufwenberg, M., & Gneezy, U. (2005). Gender & Coordination. In R. Zwick & A. Rapoport, *Experimental Business Research* (pp. 253–262–262). Boston, MA: Springer US.

Durlauf, S. (1999). The case "against" social capital. *Focus 20*(3), 1–5.

Eagly, A. H. (2009). The his and hers of prosocial behavior: an examination of the social psychology of gender. *American Psychologist*, *64*(8), 644.

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.

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. Psychological bulletin, 108(2), 233.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735–754.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*(6), 408.

Eckel, C. C., & Grossman, P. J. (1998). Are Women Less Selfish Than Men?: Evidence From Dictator Experiments. *The Economic Journal*, *108*(448), 726–735.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry*, *39*(2), 171–188.

Eckel, C. C., & Grossman, P. J. (2008a). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In C. R. Plott & V. L. Smith, *Handbook of experimental economics results* (Vol. 1, pp. 509–519). Elsevier

Eckel, C. C., & Grossman, P. J. (2008b). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1–17.

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *The American Economic Review*, *88*(4), 848–881

Fiddick, L., Cosmides, L., Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason Selection Task. *Cognition* 75: 1-79.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology. 82*(6), 878-902.

Francis, B., Hasan, I., Park, J. & Wu, Q. (2014). Gender differences in Financial Reporting Decision-Making: Evidence from Accounting Conservatism. *Contemporary Accounting Research,* in press.

Francis, B., Hasan, I., & Wu, Q. (2012). The Impact of CFO Gender on Bank Loan Contracting. *Journal of Accounting, Auditing & Finance*, *28*(1), 53–78.

Francis, J., LaFond, R., Olsson, P., & Schipper, K. (2005). The market pricing of accruals quality. *Journal of Accounting and Economics* 39, 295–327.

Fryer, R. G., Levitt, S. D., & List, J. A. (2008). Exploring the Impact of Financial Incentives On Stereotype Threat: Evidence from a Pilot Study. *The American Economic Review*, 98(2), 370–37

Fudenberg, D., & Levine, D. K. (2009). Learning and equilibrium. *Annual Review of Economics*, *1*(1), 385-420.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences.* American Psychological Association.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... & López, W. L. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of personality and social psychology*, *79*(5), 763.

Goette, L., Huffman, D., & Meier, S. (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics*, *4*(1), 101-115.

Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review*, *95*(1), 384–394.

Gold, A., Hunton, J. E., & Gomaa, M. I. (2009). The Impact of Client and Auditor Gender on Auditors' Judgments. *Accounting Horizons*, *23*(1), 1–18.

Grafen A (1990) Biological signals as handicaps. *J Theor Biol* 144(4):517–546

Gray, J. (1992). Men are from Mars, women are from Venus: A practical guide for improving communication and getting what you want in your relationships. New York: HarperCollins.

Gul, F. A., Fung, S. Y. K., & Jaggi, B. (2009). Earnings quality: Some evidence on the role of auditor tenure and auditors' industry expertise. *Journal of Accounting and Economics*, *47*(3), 265–287.

Gul, F. A., Hutchinson, M., & Lai, K. M. Y. (2013). Gender-Diverse Boards and Properties of Analyst Earnings Forecasts. *Accounting Horizons*, *27*(3), 511–538.

Gul, F. A., Srinidhi, B., & Ng, A. C. (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics*, *51*(3), 314–338.

Gul, F. A., Srinidhi, B., & Tsui, J. S. L. (2008). Board Diversity and the Demand for Higher Audit Effort. *Working paper.*

Hardies, K., Breesch, D. & Branson., J. (2014). Do (fe)male auditors impair audit quality? Evidence from going-concern opinions. *European Accounting Review,* in press.

Hargreaves-Heap, S. P., & Zizzo, D. J. **(**2009). The Value of Groups. *American Economic Review*, *99*(1), 295–323.

Hess NC, Hagen EH. (2006) Sex differences in indirect aggression: Psychological evidence from young adults. *Evolution and Human Behavior.*

Holm, H. (2010) Truth and lie detection in bluffing, *Journal of Economic Behavior & Organization*, 76(2): 318-324.

Hooks, K. L. (1992). Gender effects and labor supply in public accounting: An agenda of research issues. *Accounting, Organizations and Society*, *17*(3–4), 343–366.

Isaac, C., Lee, B., & Carnes, M. (2009). Interventions that affect gender bias in hiring: a systematic review. *Academic Medicine. 84*(1), 1440-1446.

Ittonen, K., Miettinen, J., & Vähämaa, S. (2010). Does Female Representation on Audit Committees Affect Audit Fees? *Quarterly Journal of Finance and Accounting*, *49*(3/4), 113–139.

Ittonen, K., & Peni, E. (2011). Auditor's Gender and Audit Fees. *International Journal of Auditing,* 16(1), 1–18

Ittonen, K., Vähämaa, E., & Vähämaa, S. (2013). Female Auditors and Accruals Quality. *Accounting Horizons*, *27*(2), 205–228.

Johnson, E., Kaplan, S., & Reckers, P. (1998). An examination of potential gender-based differences in audit managers' performance evaluation judgments. *Behavioral Research in Accounting*, *10*, 47–75.

Jones, K. L., Krishnan, G. V., & Melendrez, K. D. (2008). Do Models of Discretionary Accruals Detect Actual Cases of Fraudulent and Restated Earnings? An Empirical Analysis. *Contemporary Accounting Research*, *25*(2), 499–531.

Jost, J.T. and M.R. Banaji, 1994. "The role of stereotyping in system-justification and the production of false consciousness." *British Journal of Social Psychology* 33(1), 1–27.

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. *Handbook of prejudice, stereotyping, and discrimination*, 199-227.

Kawakami, K., Dovidio, J.F., & van Kamp, S. (2005). Kicking the habit: Effects of nonstereotypic association training and correction processes on hiring decisions. *Journal of Experimental Social Psychology. 41*, 68-75.

Kenrick, D. T. (1987) Gender, genes, and the social environment: A biosocialinteractionist perspective. In: *Review of Personality and Social Psychology, vol. 7,* ed. P. Shaver & C. Hendrick. Sage.

Kornberger, M., Carter, C., & Ross-Smith, A. (2010). Changing gender domination in a Big Four accounting firm: Flexibility, performance and client service in practice. *Accounting, Organizations and Society*.

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics*, *39*(1), 163–197.

Krishnan, G. V., & Parsons, L. M. (2007). Getting to the Bottom Line: An Exploration of Gender and Earnings Quality. *Journal of Business Ethics*, *78*(1-2), 65–76.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America. 98*(26), 15387-15392.

Levinson, D. (1989) *Family violence in cross-cultural perspective*. Sage.

Lewis DK (1969) *Convention: A Philosophical Study* Harvard Univ Press, Cambridge, MA.

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., et al. (2014). Gender Stereotypes of Personality: Universal and Accurate? *Journal of Cross-Cultural Psychology*, 45(5), 675–694.

Marlowe FW. Hunting and gathering: the human sexual division of foraging labor. *Cross-Cultural Research.* 2007;41(2):170–195.

Martin, A. D., Nishikawa, T., & Williams, M. A. (2009). CEO Gender: Effects on Valuation and Risk. *Quarterly Journal of Finance and Accounting*, *48*(3), 23–40.

New, J., Krasnow, M. M., Truxaw, D., & Gaulin, S. J. (2007). Spatial adaptations for plant foraging: women excel and calories count. *Proceedings of the Royal Society B: Biological Sciences, 27*4(1626), 2679-2684.

Niskanen, J., Karjalainen, J., Niskanen, M., & Karjalainen, J. (2011). Auditor gender and corporate earnings management behavior in private Finnish firms. *Managerial Auditing Journal, 26*(9), 778-793.

Numan, W., & Willekens, M. (2012). An empirical test of spatial competition in the audit market. *Journal of Accounting and Economics*, *53*(1–2), 450–465.

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334

O'Donnell, E. F., & Johnson, E. N. (2001). The Effects of Auditor Gender and Task Complexity on Information Processing Efficiency. *International Journal of Auditing*, *5*(2), 91–105.

Orbell, J., Dawes, R., & Schwartz-Shea, P. (1994). Trust, social categories, and individuals: The case of gender. *Motivation and Emotion*, *18*(2), 109–128.

Ortmann, A., & Tichy, L. K. (1999). Gender differences in the laboratory: evidence from prisoner's dilemma games. *Journal of Economic Behavior & Organization*, *39*(3), 327–339.

Otterbein, K. F., & Otterbein, C. S. (1965). An Eye for an Eye, A Tooth for a Tooth: A Cross-Cultural Study of Feuding. *American Anthropologist, 67*(6), 1470-1482.

Otterbein, K. F. (1968). Internal War: A Cross-Cultural Study. *American Anthropologist*, 70(2), 277-289.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, *62*(4), 659–661.

Pickering, M., 2001. *Stereotyping: The politics of Representation.* Palgrave.

Posner, M. I., & Konick, A. F. (1966). Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance,* 1(1), 71-86.

Pruitt, D. G. (1983). Strategic Choice in Negotiation. *American Behavioral Scientist*, *27*(2), 167–194.

Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters*, 123(1), 33–36.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12), 4403-4408.

Rosenkrantz, P., Vogel, S., Bee, H., Borverman, I., & Broverman, D. M. (1968). Self-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology. 32,* 287-295.

Sanchez-Pages, S., & Vorsatz, M. (2007). An experimental study of truth-telling in a sender-receiver game. *Games and Economic Behavior*, *61*(1), 86–112.

Savage, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, 66(336), 783–801.

Schniter, E., & Shields, T. W. (2014). Ageism, honesty, and trust. *Journal of Behavioral and Experimental Economics,* 51(0), 19–29.

Schniter, E., Sheremeta, R. M., & Sznycer, D. (2013). Building and rebuilding trust with promises and apologies. *Journal of Economic Behavior & Organization, 94*, 242-256.

Searcy WA, Nowicki S (2005) *The Evolution of Animal Communication*. Princeton Univ Press, Princeton.

Shelley, G. P., Page, M., & Kuhlman, D. M. (2010). Semantic Differential Judgments of Cooperators, Competitors and Individualists at Zero-Acquaintance. *Letters on Evolutionary Behavioral Science*, *1*(1).

Sidanius, J., & Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. New York: Cambridge University Press.

Smuts, B. B., Cheney, D. L., Seyfarth, R. M., & Wrangler, R. W. (1987). *Primate Societies*. University of Chicago Press.

Solnick, S. J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*(2), 189–200.

Spence M (1973) Job market signaling. *Quarterly Journal of Economics,* 87(3):355–374.

Srinidhi, B., Gul, F. A., & Tsui, J. (2011). Female Directors and Earnings Quality. *Contemporary Accounting Research*, *28*(5), 1610–1644.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences*, 99(17), 11531-11536.

Sun, J., Liu, G., & Lan, G. (2011). Does Female Directorship on Independent Audit Committees Constrain Earnings Management? *Journal of Business Ethics*, *99*(3), 369–382.

Sutter, M. (2009). Individual Behavior and Group Membership: Comment. *The American Economic Review*, *99*(5), 2247–2257.

Sutter, M., Bosman, R., Kocher, M. G., & Winden, F. (2009). Gender pairing and bargaining—Beware the same sex! *Experimental Economics*, *12*(3), 318–331.

Thiruvadi, S., & Huang, H.-W. (2011). Audit committee gender differences and earnings management. *Gender in Management: an International Journal*, *26*(7), 483–498.

Tajfel, H. (1981). Social stereotypes and social groups. In J. C. Turner & H. Giles (Eds.). *Intergroup Behaviour* (pp. 144-167). Oxford: Blackwell.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Cambell (Ed.),Sexual selection and the descent of man, 1871–1971 (pp. 136–179). London:Heinemann

Vaillancourt, T., & Sharma, A. (2011). Intolerance of sexy peers: Intrasexual competition among women. *Aggressive behavior*, *37*(6), 569-577.

Van Velzen, H. U. E. T., & Van Wetering, W. (1960). Residence, power groups and intra-societal aggression. *International Archives of Ethnography, 49*, 169-200.

Van Vugt, M. (2011). The male warrior hypothesis. *The Psychology of Social Conflict and Aggression, 1*3, 233.

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition the Male-Warrior hypothesis. *Psychological Science, 18*(1), 19-23.

Van Vugt, M., & Park, J. H. (2009). Guns, germs, and sex: how evolution shaped our intergroup psychology. *Social and Personality Psychology Compass*, *3*(6), 927-938.

Waller, W. S., & Felix, W. L., Jr. (1984a). The Auditor and Learning From Experience: Some Conjectures. *Accounting, Organizations and Society*, 9(3/4), 383–406.

Waller, W. S., & Felix, W. L., Jr. (1984b). The effects of incomplete outcome feedback on auditors' self-perceptions of judgment ability. *Accounting Review*, 59(4), 637–64

Walters, A. E., Stuhlmacher, A. F., & Meyer, L. L. (1998). Gender and Negotiator Competitiveness: A Meta-analysis. *Organizational Behavior and Human Decision Processes*, *76*(1), 1–29.

Williams, J. E., & Best, D. L. (1982). *Measuring sex stereotypes: A thirty-nation study.* Beverly Hills, CA: Sage.

Windsor, C., & Auyeung, P. (n.d.). The effect of gender and dependent children on professional accountants' career progression. *Critical Perspectives on Accounting*, *17*, 828–844.

Winkler, R. L., & Murphy, A. H. (1968). "Good" Probability Assessors. *Journal of Applied Meteorology*, *7*(5), 751–758.

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin*, *128*, 699-727.

World Bank (2007). *World Development Indicators*. World Bank, Washington D.C.

Zahavi, A. (1975). Mate selection-a selection for a handicap. *Journal of Theoretical Biology,* 53(1): 205–214.

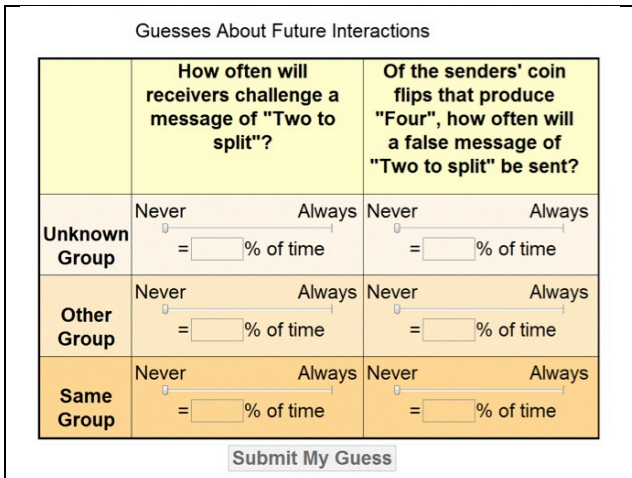
Zimbelman, M. F., & Waller, W. S. (1999). An Experimental Investigation of Auditor-Auditee Interaction Under Ambiguity. *Journal of Accounting Research*, 37, 135–155.

# 7 Appendix

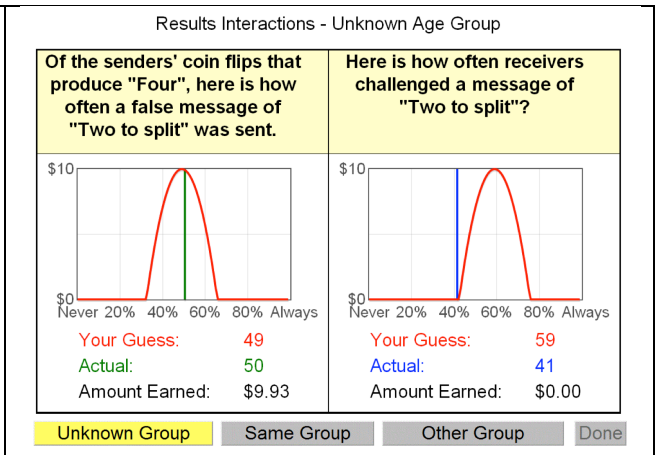
## 7.1 Instructions

Prerecorded multimedia instructions consisting of video and audio were projected onto a large screen viewable to all and played back over a public address system. The multimedia file has been attached separately.

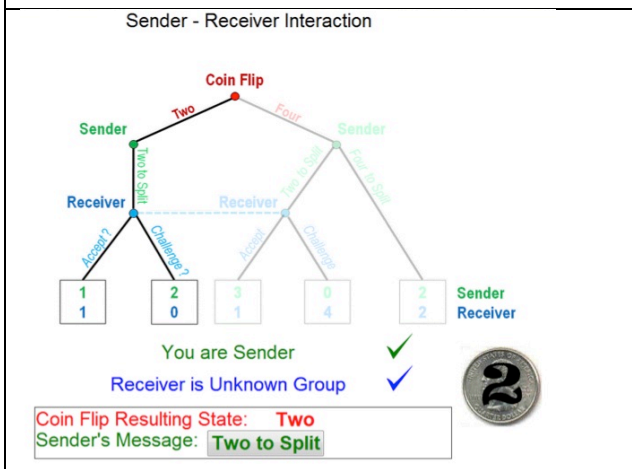
## 7.2 Screenshots



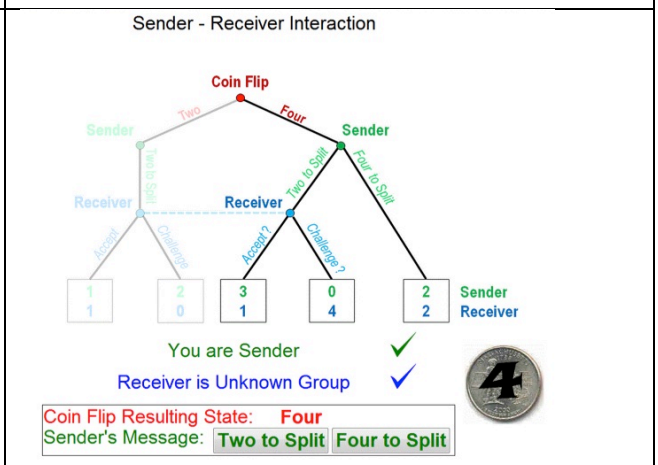

(i) Input screen for guesses about future interactions.



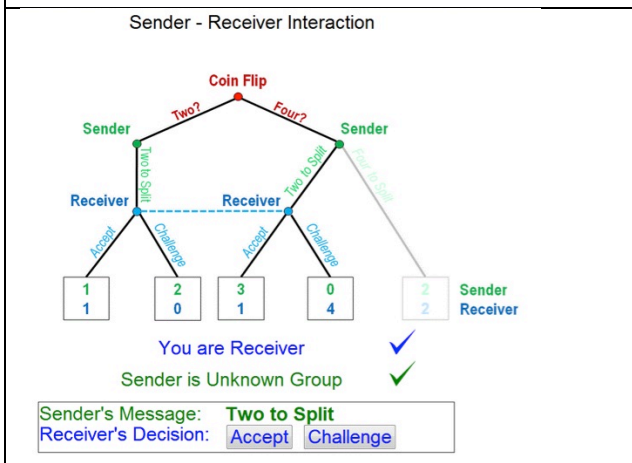

(ii) Feedback screen showing actual behavior and guesses scored according to quadratic scoring rule.



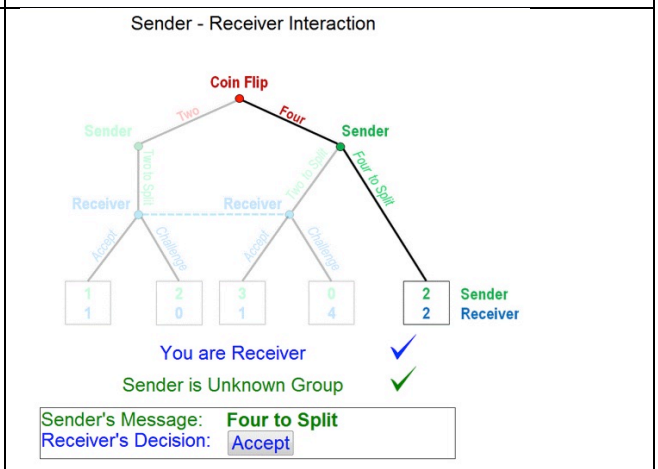

(iii) Issuer screen for coin flip of heads (= Two).




(iv) Issuer screen for coin flip of tails (= Four).




(v) Auditor screen for "Two to split" message.




(vi) Auditor screen for "Four to split" message.