ABOUT PSI CHI
Psi Chi is the International Honor Society in Psychology, founded in 1929 for the purposes of encouraging, stimulating, and maintaining excellence in scholarship, and advancing the science of psychology. Membership is open to graduate and undergraduate students making the study of psychology one of their major interests and who meet Psi Chi’s minimum qualifications. Full-time faculty with a master’s or doctorate in psychology or a closely related field are also invited to join. Psi Chi is a member of the Association of College Honor Societies (ACHS), and is an affiliate of the American Psychological Association (APA) and the Association for Psychological Science (APS). Psi Chi’s sister honor society is Psi Beta, the national honor society in psychology for community and junior colleges.

Psi Chi functions as a federation of chapters located at over 1,100 senior colleges and universities around the world. The Psi Chi Central Office is located in Chattanooga, Tennessee. A Board of Directors, composed of psychology faculty who are Psi Chi members and who are elected by the chapters, guides the affairs of the Organization and sets policy with the approval of the chapters.

Psi Chi serves two major goals. The first of these is the Society’s obligation to provide academic recognition to its inductees by the mere fact of membership. The second is the opportunity of each of the Society’s local chapters to nourish and stimulate the professional growth of all members through fellowship and activities designed to augment and enhance the regular curriculum. In addition, the Organization provides programs to help achieve these goals including conventions, research awards and grants competitions, and publication opportunities.

JOURNAL PURPOSE STATEMENT
The twofold purpose of the *Psi Chi Journal of Psychological Research* is to foster and reward the scholarly efforts of psychology students as well as to provide them with a valuable learning experience. The articles published in the *Journal* represent the work of undergraduates, graduate students, and faculty. To further support authors and enhance journal visibility, articles are now available in the PsycINFO® and EBSCO Academic Search Complete® databases.

JOURNAL INFORMATION
The *Psi Chi Journal of Psychological Research* (ISSN 2164-8204) is published quarterly in one volume per year by Psi Chi, Inc., The International Honor Society in Psychology, P.O. Box 709, Chattanooga, TN 37401-0709.

Subscriptions are available on a calendar-year basis only (Spring–Winter). U.S. rates are as follows (four issues): Individual $20; Institution $40.

For international rates or other information, contact Psi Chi Central Office. Periodicals postage paid at Chattanooga, TN, and additional mailing offices.

Statements of fact or opinion are the responsibility of the authors alone and do not imply an opinion on the part of the officers or members of Psi Chi. Psi Chi does not accept paid advertising for *Psi Chi Journal of Psychological Research*.

Postmaster: Send address changes to the address below.

Copyright 2015 by Psi Chi, The International Honor Society in Psychology.

For more information, contact:
Psi Chi Central Office
Publication and Subscriptions
P.O. Box 709
Chattanooga, TN 37401-0709
(423) 756-2044
www.psichi.org
psichjournal@psichi.org
The Effects of Stereotype Threat and Power on Women’s and Men’s Outcomes in Face-to-Face and E-mail Negotiations
Jasmine McCormick and Wendy L. Morris
McDaniel College

Do We Have Fun When Time Flies?
Matthew Pageau, Saint Joseph’s University
Seth Surgan, Worcester State University

Think About It: Cognitive Load and the Trolley Scenario as an Analogue of Gun Violence
Molly Ramsden
Castleton State College

Is the Relationship Between Anxiety and Creativity Moderated by Other Emotional States?
Glenn B. Little and Karl Wuensch
East Carolina University

Effects of Career-Oriented Information on Romantic Desirability
Amanda L. Young
Lenoir-Rhyne University

It’s Just a Joke: Reactions to and Justifications for Gender Role Stereotypes in Advertisements
Natalie Peters, Lucie Holmgreen, and Debra Oswald
Marquette University

Internal Martial Arts Training and the Reduction of Hostility and Aggression in Martial Arts Students
James Hernandez and Kathryn B. Anderson
Our Lady of the Lake University
Sex differences in salary negotiation have been suggested as one of the causes of the “glass-ceiling,” the pattern that women do not advance as far in their careers as men do (Stuhlmacher & Walters, 1999). Men are four times more likely than women to negotiate for their salary and, when it comes to first job offers, only 7% of women negotiate whereas 58% of men do (Babcock & Laschever, 2003). According to a meta-analysis (Stuhlmacher & Walters, 1999), when women do negotiate, they negotiate for lower profits than men. These sex differences in negotiation are particularly problematic given that the amount of money asked for during a negotiation largely determines the outcome salary (Barron, 2003).

The purpose of the current research was to better understand multiple factors that may increase or decrease gender differences in negotiation outcomes. Because women are stereotyped as being less effective negotiators than men (Kray, Thompson, & Galinsky, 2001), they tend to negotiate less successfully than men when they experience stereotype threat (Kray et al., 2001; Tellhed & Björklund, 2011). The current research examined situations that may minimize the effects of stereotype threat and reduce the salary gap. Negotiation outcomes were explored in situations in which the salience of sex varies (face-to-face vs. e-mail negotiations and same vs. mixed sex negotiations) and in situations in which sex differences were likely to be minimized when women were in a position of power.
Prescriptive Stereotypes
Women may negotiate less frequently and less aggressively than men due to the influence of prescriptive stereotypes regarding sex (Bowles & Babcock, 2009). Stereotypes not only provide a description of what people think men and women are like, but also a prescription of how men and women ought to behave in order to fulfill expectations of their gender (Burgess & Borgida, 1999; Eagly, 1987). Stereotypical behaviors are expected, and negative social consequences can occur if they are not followed (Bowles, 2012). Trying to bargain for more money for oneself such as during a salary negotiation is against the feminine role, and women are seen as less persuasive and less likeable when they violate that role (von Hippel, Wiryakusuma, Bowden, & Shochet, 2011).

To be a successful negotiator, people think they must focus on the competitive aspects of negotiation rather than the cooperative aspects (Thompson, Valley, & Kramer, 1995). Successful negotiation is associated with stereotypical masculine traits such as being assertive and decisive, and unsuccessful negotiation is associated with stereotypical feminine traits such as being emotional and accommodating (Kray et al., 2001). Thompson et al. (1995) found that feelings of success following a negotiation were “bittersweet” because successful negotiators felt less friendly, cooperative, and sincere. The results of these studies suggested that success and competition in negotiation are perceived as being in direct conflict with positive stereotypes associated with femininity. Thus, women may not negotiate as frequently or as successfully as men due to an implicit desire to adhere to their sex role.

Although it is possible that women are not as skilled as men at negotiation, evidence has shown that this is not the case. For example, when women are informed that stereotypically feminine traits are linked to successful negotiation, their performance improves (Kray, Galinsky, & Thompson, 2002). Although prescriptive stereotypes depict women as lacking assertiveness in advocating for themselves, it is acceptable for them to be assertive in a selfless act or on someone else’s behalf because this behavior is consistent with communal stereotypes of women (Amanatullah & Morris, 2010). In fact, when women negotiate on someone else’s behalf, they negotiate more effectively, and sex differences in negotiation outcomes decreases (Amanatullah & Morris, 2010; Bowles, Babcock, & McGinn, 2005). Thus, women are quite capable of negotiating successfully in situations that do not conflict with the prescriptive stereotypes for women.

Stereotype Threat
Because successful performance in negotiation is associated with masculine traits such as rationality and assertiveness, and poor performance in negotiation is associated with feminine traits such as compassion and unassertiveness, people expect women to do poorly in negotiation, thus making them vulnerable to stereotype threat (Kray et al., 2001). Stereotype threat is a social-psychological threat that one may confirm a negative stereotype about one’s group while performing a specific task, and this threat can lead to underperformance (Steele, 1997). Both Kray et al. (2001) and Tellhed and Björklund (2011) have demonstrated that women underperform in hypothetical salary negotiations when stereotype threat is activated. When participants were told that the experimental task was diagnostic of their ability to negotiate, women negotiated less effectively than men. However, when they were told that the task was not diagnostic of their ability to negotiate, the sex difference decreased.

Research on stereotype threat has consistently found that stereotype threat not only affects the group which is stereotyped to do poorly, but also the group stereotyped to do well (Walton & Cohen, 2003). Groups who are perceived as being advantaged at a particular task tend to perform better when stereotype threat is activated (Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999). This finding is referred to as stereotype boost or stereotype lift, and has been supported by meta-analytic findings (Walton & Cohen, 2003). Both the positive and negative effects of stereotype threat are influenced by various situational factors.

Situational Factors That May Impact the Effects of Stereotypes
Sex composition of the negotiating pair. Stereotype threat is especially likely to affect behavior in situations in which the participant’s sex is salient (McGlone & Aronson, 2006). People are more likely to think about their own sex when they are in mixed-sex interactions than in same-sex interactions (Cota & Dion, 1986). Because sex is more salient in these mixed-sex interactions, gender roles, prescriptive stereotypes, and stereotype threat may have a stronger influence on
behaviors in these situations. For example, women do worse in a problem-solving task when in the presence of men than when in the presence of women (Inzlicht & Ben-Zeev, 2000). Although past research has explored the effects of sex composition on negotiation (Kray & Thompson, 2005), the question of whether stereotype threat affects negotiation outcomes in same-sex dyads differently than mixed-sex dyads has not been explored. The effects of stereotype threat are likely to be greater in mixed-sex negotiations where sex is more salient than in same-sex negotiations.

**Power.** Although sex differences in negotiation are likely to be greater in mixed-sex negotiations than in same-sex negotiations (Kray et al., 2001), Watson (1994) argued that the reason men are likely to perform better than women in mixed-sex negotiations is that they are assumed to have more power and that power is actually a better predictor than sex of negotiation outcome. Sex stereotypes are associated with cues related to power and status (Glick & Fiske, 2001; Ridgeway, 2011). Stereotypically feminine communal traits such as helping others, being unassertive, and being selfless are consistent with expectations of low-power behavior. In fact, a prescriptive stereotype of both women and low-power people is to be accommodating of other people’s needs, which makes them less threatening (Jackman, 1994; Ridgeway, 2011). On the other hand, stereotypically agentic masculine traits such as being competitive and aggressive are associated with high-power behavior (Conway, Pizzamiglio, & Mount, 1996). Because sex serves as a cue for power, interactions involving mixed-sex pairs often carry the assumption that the man holds more power than the woman (Eagly, 1983; Wood & Karten, 1986). A difficulty in separating sex effects from power effects in mixed-sex interactions is that sex is often confounded with perceptions of power, and perhaps power, rather than gender per se, is responsible for sex differences in negotiation outcomes.

Watson (1994) found that power is a better predictor of negotiation outcomes and success than sex. When men and women were randomly assigned to high or low power positions in same- or mixed-sex dyads, power consistently predicted negotiation outcomes more so than sex did. Although power is a strong predictor of negotiation outcomes, Watson (1994) also found that sex effects emerged in mixed-sex dyads when the two people were of similar power status, presumably because men were assumed to have more power than women in those interactions.

Studies that have examined the effects of stereotype threat on sex differences in negotiation have not typically manipulated power. To disentangle the effects of sex from the effects of power, it is important to manipulate whether women or men are in the more powerful or less powerful role and also whether the negotiating dyads consist of same-sex or mixed-sex pairs. It is quite possible that the sex differences in negotiation outcomes caused by stereotype threat may be minimized when women are randomly assigned to a position of high power and men are assigned to a position of low power. Similarly, past research on social role theory (Eagly, 1987) has demonstrated that sex differences in verbal communication behaviors are minimized when power is manipulated in experiments (Johnson, 1994).

**Mode of communication.** Just as power may reduce sex differences, so too may virtual forms of communication such as e-mail in which sex is less salient than in face-to-face interactions. Although meta-analytic findings have shown that women typically behave more cooperatively than men during negotiation, a behavior that is not conducive to success, this sex difference disappears when the negotiation is not a face-to-face interaction (Walters, Stuhlmacher, & Meyer, 1998).

According to Mischel (1977), certain situations are considered “weak situations” because there are no clear expectations regarding how one is supposed to behave. Weak situations allow people to act in idiosyncratic ways rather than being constrained by certain social roles such as sex. E-mail is considered a weak situation given the weakened social norms, the decreased salience of sex, and ambiguity due to a lack of nonverbal cues (Sproull, Subramani, Kiesler, Walker, & Waters, 1996; Stuhlmacher, Citera, & Willis, 2007). E-mail has been associated with an increase in counter-normative behavior (Kiesler & Sproull, 1992). When people negotiate by e-mail, they feel less constrained by social norms and are more likely to make threats and issue ultimatums (Morris, Nadler, Kurtzberg, & Thompson, 2000). Women in particular behave more aggressively in virtual negotiations than in face-to-face negotiations (Stuhlmacher et al., 2007), and this could potentially improve their negotiation outcomes.

Given the weak nature of e-mail and the decreased salience of sex, it is possible that the effects of stereotypes threat may be smaller in e-mail negotiations than in face-to-face negotiations.
Furthermore, if it is actually power, rather than sex, that affects the negotiation outcome, e-mail negotiations may still provide a more equitable mode of negotiation given that the impact of both power and sex roles is lessened in virtual communication (Sproull & Kiesler, 1986). Given the ubiquitous use of various forms of electronic communication, it is likely that the frequency of virtual negotiation will only increase in years to come. Studying the impact of e-mail negotiation will serve to deepen the understanding of factors that may decrease the negative effects of stereotype threat and reduce the sex gap in salary negotiations.

The Current Research
The negotiation outcomes of men and women were tested in various conditions predicted to strengthen or weaken the effects of stereotype threat. Participants were randomly assigned to negotiate with or without stereotype threat activated in situations in which sex was more salient (mixed-sex negotiations) or less salient (same-sex negotiations). In order to separate the effects of power, men and women were randomly assigned to a position of high power (employer) or low power (employee) during the negotiation. Approximately half of these negotiations took place in a face-to-face interaction and the others took place via e-mail to determine if e-mail negotiations were resistant to the effects of stereotype threat.

The experiment tested three primary hypotheses. Consistent with past research on stereotype threat, we hypothesized that women would negotiate less effectively when under stereotype threat than when not under stereotype threat. On the other hand, we expected men to experience a boost under conditions of stereotype threat that would improve their negotiation performance. Specifically, these effects of stereotype threat (negative for women and positive for men) were predicted to be strongest in situations in which sex was most salient such as in mixed-sex pairs. Consistent with past research that power is the strongest determinant of negotiation outcomes, we also predicted that power would be a stronger predictor of outcomes than sex or stereotype threat. Third, because e-mail negotiations are expected to be a relatively weak situation where sex would be less salient, we predicted that women would negotiate more effectively through e-mail than face-to-face, particularly when under conditions of stereotype threat. In e-mail negotiations, we expected that women would be less constrained by prescriptive stereotypes and would negotiate as effectively as men.

Method
Participants
Although 110 people (55 pairs) participated in the present experiment, seven pairs were dropped from the data analysis because they did not reach an agreement in the allotted time and therefore did not produce any financial data on the outcome of the negotiation. Ninety-six participants (48 pairs, 59 women, 37 men) were included in the data analysis. The age range of participants was 18 to 25 with a median age of 19. Seventy-five percent of participants were European American, 13% were African American, 7% were Asian, and 5% were Latino. The experiment was advertised via e-mails sent out to the entire college community and posted on a website for students taking an introductory psychology course in which students were required to learn about research by various methods including participating in experiments. Participants taking the introductory psychology course earned credits by participating. All participants were also eligible to win one of four $100 lottery prizes offered as an incentive. In planning each experimental session, we attempted to pair together participants who were not already friends. A questionnaire given to participants at the end of each session indicated that 72% did not know their partner at all, 24% did not know their partner very well, and only 4% knew their partner fairly well. The college’s institutional review board approved this study.

Design
This was a 2 x 2 x 2 x 2 x 2 between participants design. The independent variables included the sex of the participant, whether the participant was in a position of power (employer or employee), the sex composition of the dyad (same-sex or mixed-sex pairs), the mode of communication (face-to-face or e-mail), and stereotype threat (present or absent). The dependent variables were the monetary outcome of the salary negotiation and the expected outcome of the negotiation. Although the topic of the negotiation was about a hypothetical salary, a small amount of real money was at stake in the negotiation as an incentive for the participants to take the negotiation seriously. Ten dollars was allocated to each pair and the amount that each individual in the pair earned of the $10 was dependent upon how well he or she negotiated.
Stereotype Threat and Negotiation | McCormick and Morris

Materials

**Informational packet.** Each participant was given an informational packet, which included a description of the hypothetical job for which a salary would be negotiated, the guidelines for the negotiation, a salary goal sheet, and a paragraph manipulating stereotype threat.

**Job description.** Participants were given a fictional job advertisement for a position of facility manager at a pharmaceutical company. The job ad described the company, the responsibilities of the position, and entry requirements for the position. Participants in the employee role were told that they should imagine that they had graduated college, been offered this position, and now needed to negotiate their starting salary. Participants in the employer role were told that they should imagine that they had been approached by Tellhed and Björklund (2011).

**Guidelines for the negotiation.** The informational packet described the guidelines for the negotiation including the role the participant would play (employer or employee), how the negotiation would occur (face-to-face or via e-mail), and that there would be a 15 min time limit. The guidelines stated that participants could create new facts about their character during the negotiation and bring in information beyond what they had been given in the packet.

**Salary goal sheet.** The salary goal sheet included a chart with 11 salary ranges that the pair could agree upon from $14,000 to $36,000. The employers and employees received different salary goal sheets, listing their instructed goal in the negotiation and the actual payout they would receive depending upon how well they did in the negotiation. The instructed goal for employers was $17,000, and the highest salary they were told to accept was $31,000. The instructed goal for employees was $33,000 and the lowest salary they were told to accept was $19,000. Each salary range on the chart was listed with an actual financial reward that participants could earn and provided an incentive to do well in the negotiation. For example, if an employer managed to reach the goal of a $17,000 salary, that participant would earn $9 and the employee would only earn $1. Whereas if an employee reached the goal of $33,000, that participant would earn $9 and the employer would only earn $1. Ten dollars was allotted to each pair and the outcome of the negotiation determined how that money was allocated between the two negotiation partners ($10-$0, $9-$1, $8-$2, $7-$3, $6-$4, $5-$5). The $10 was used to motivate the participants to do well in the negotiation and to increase the realism of the situation by providing varying financial rewards for degrees of success in negotiating (as used in previous research by Bowles et al., 2005). The informational packet stated that no money would be given to either participant if no agreement was reached.

**Stereotype threat manipulation.** Stereotype threat was manipulated in the informational packet via a paragraph that either increased stereotype threat or minimized it by describing the negotiation task as either challenging and diagnostic of their skills or as easy for novices and not an assessment of their skills. These same two paragraphs were used as a stereotype threat manipulation in a previous experiment about negotiation (Tellhed & Björklund, 2011). In each pair, both participants were given the same paragraph either activating or minimizing stereotype threat.

**Prenegotiation survey.** This survey included demographic questions about the participants and served as a manipulation check to verify that the participants understood their role during the negotiation (i.e., employer or employee), what their instructed goal salary should be, and what their lowest/highest acceptable salary should be. Participants also predicted what salary would be the outcome of the negotiation.

Procedure

Pairs of participants were randomly assigned to conduct a face-to-face or virtual negotiation condition with stereotype threat present or absent. After participants met each other and signed the consent form, they were randomly assigned to the roles of employer and employee, and then read the informational packet and completed the prenegotiation survey alone in separate rooms. Afterward, participants in the face-to-face condition were brought together in a room with a table, two chairs on opposite sides, and a digital camcorder to record the negotiation process. In the virtual condition, participants stayed in separate rooms to conduct the negotiation using computers and Gmail accounts created for the purpose of this experiment. The experimenter was not present during the negotiations but did step in to alert the participants when only 5 min remained. At the end of 15 min of negotiation or whenever an agreement...
was reached, the participants reported the outcome of the negotiation and were paid their portion of the $10. If no agreement was reached, participants received no money. Upon completion, participants were debriefed and entered into a lottery to win several $100 prizes.

**Results**

This 2 x 2 x 2 x 2 x 2 between-participants design was analyzed using generalized estimating equations (GEE) to account for the fact that the negotiation outcomes of people from the same pair were correlated with each other (i.e., if someone did very well, his or her negotiation partner could not have done very well). Thus individuals were matched with their negotiation partner in all analyses. The independent variables included the sex of the participant, whether the participant was in a position of power (employer or employee), the sex composition of the dyad (same-sex or mixed-sex pairs), the mode of communication (face-to-face or e-mail), and stereotype threat (present or absent).

The dependent variables were the financial outcome of the negotiation (how much better or worse they did than the goal) and the expected outcome (how much better or worse they expected to do than the goal). For employers, doing better would mean reaching an agreement with a lower salary, whereas doing worse would mean reaching an agreement with a higher salary. The opposite was true for employees because they wanted to receive as much money as possible from the employer; doing better would mean reaching an agreement with a higher salary. Because most participants did not actually reach their instructed goal, in all of the results presented below, the negative amounts of money represent how much worse the participant did than the goal or how much worse they expected to do than the goal.

GEE is an extension of generalized linear models, which is preferred over within-participants Analysis of Variance (ANOVA) because an ANOVA does not acknowledge the covariance that occurs within repeated-measures or nested designs (Ballingner, 2004). Generalized linear models instead assume independence of observations. Because the design of the present study included pairs of participants whose outcomes were dependent upon each other, GEE was used for this nested design. Horton and Lipsitz (1999) have warned that GEE results should be interpreted with caution when there are fewer than 20 clusters of correlated data. In the present study, there were 48 clusters or pairs of correlated data. Although this sample size was sufficient to reject the null hypothesis for the main effects (see Table 1), nonsignificant effects for the higher order interactions must be interpreted with caution due to the small number of dyads per condition (see Table 2).

**The Effects of Stereotype Threat in Mixed-Sex Dyads**

There were no significant main effects of stereotype threat, Wald Chi-Square = .05, p = .82, d = 0.02, dyad type, Wald Chi-Square = .05, p = .82, d = 0.02, or sex, Wald Chi-Square = 2.81, p = .09, d = 0.17, on the outcome of the negotiation. However, there was a three-way interaction between stereotype threat, sex, and dyad type on the outcome of the negotiation, indicating that men got a positive boost from stereotype threat and women did worse under stereotype threat; these effects only occurred in mixed-sex dyads, Wald Chi-Square = 9.85, p = .002. When in mixed-sex dyads, men came $3,756 closer to the goal salary when under stereotype threat than without threat, p = .003, d = 0.30, 95% CI [$1,265, $6,247], whereas women were $3,756 further from their goal salary when under stereotype threat than without threat, p = .003, d = 0.30, 95% CI [$1,265, $6,247]. See Figure 1. However, when in same-sex dyads, there was no significant difference in the outcome of the negotiation between the stereotype threat condition and the no-threat threat condition for men, p = .53, or for women, p = .80.

**The Effect of Power on Outcomes and Expectations**

There was a main effect of power on how close participants got to their goal, Wald Chi-Square = 22.74, p < .001, d = 0.49. Participants in the role of the employer got closer to their goal ($M = -$9,608, $SD = $4,802) than participants in the role of employee ($M = -$9,608, $SD = $4,998).

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Participants in Each Condition for the Main Effect Comparisons</strong></td>
</tr>
<tr>
<td>Independent variable</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>Stereotype threat</td>
</tr>
<tr>
<td>Mode of communication</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Dyad type</td>
</tr>
</tbody>
</table>
by an average amount of $3,374, 95% CI [$1,987, $4,761]. There was also a main effect of power on the expected outcome of the negotiation, Wald Chi-Square = 13.05, p < .001, d = 0.37. Participants in the role of employer expected to come closer to their goal (M = $3,953, SD = $5,517) than participants in the role of employee (M = $6,789, SD = $5,361) by an average amount of $2,836, 95% CI [$1,297, $4,375]. See Figure 2. These effects of power were not qualified by any significant interactions with other independent variables. The effect of power on the outcome of the negotiation did not vary as a function of gender as indicated by the nonsignificant two-way interaction between power and gender, Wald Chi-Square = 0.58, p = .45, and power did not minimize the effects of stereotype threat as indicated by the nonsignificant two-way interaction between power and stereotype threat, Wald Chi-Square = 0.003, p = .95.

No Differences Between E-Mail and Face-to-Face Negotiations

There were no overall differences between the outcomes of the e-mail and face-to-face negotiations, Wald Chi-Square = .05, p = .82, d = 0.02. The two-way interaction between mode of communication and sex was not significant, Wald Chi-Square = 1.02, p = .31, and more specifically, it was not the case that women negotiated more successfully via e-mail than face-to-face, p = .57. Mean difference = $557, 95% CI [-$1,356, $2,470], d = 0.06. Furthermore, there was no evidence that the effect of stereotype threat was minimized in e-mail negotiations because the two-way interaction between mode of communication and stereotype threat was not significant, Wald Chi-Square = .05, p = .82.

Discussion

In support of the main hypothesis, the results indicated that the effects of stereotype threat on negotiation outcomes (positive for men and negative for women) are more likely to occur in mixed-sex dyads given the increased salience of sex. In fact, there was no evidence that the stereotype threat manipulation affected the negotiation outcomes at all in the same-sex dyad condition. However, we did not find any support for a minimized effect of stereotype threat in e-mail negotiations compared to the face-to-face negotiations. Contrary to our hypothesis and previous research (Stuhlmacher et al., 2007), the mode of communication did not affect the negotiation outcomes in this study.

As predicted, power was a strong predictor of negotiation outcomes and also expectations. People playing the role of employer expected to do better and did do better than people playing the role of employee, and the effect sizes for these results were larger than others in the present study. These results were consistent with the findings of Watson (1994) that power can have a stronger impact than sex; the effect size for sex was much smaller than the effect size for power in the present study.

Unlike other research, which has shown power to affect the behavior of men and women differently (Holleman, Cheairs, & Cook, 2013), the results of the present study did not find that power varied as a function of sex. Furthermore, although having power did improve negotiation outcomes for men and women (as the main effect of power shows), manipulating power did not decrease the effects of stereotype threat. However, the fact that power did not interact with other independent variables in the study might be due to a lack statistical power in this multifactorial design.

Limitations

Although the present study found several important significant results, it is possible that additional higher order interactions might have been statistically significant with a larger sample size. Given the five independent variables in this experiment, there were 32 possible conditions that any participant could have been in. With only 96 participants, the number of participants per condition was small, making it more difficult to find significant effects for two-way, three-way, four-way, and five-way interactions. Therefore, nonsignificant interactions

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Dyads per Condition</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Face-to-face</td>
</tr>
<tr>
<td>Stereotype threat</td>
</tr>
<tr>
<td>Same-sex male dyads</td>
</tr>
<tr>
<td>Same-sex female dyads</td>
</tr>
<tr>
<td>Mixed-sex dyad, male employer</td>
</tr>
<tr>
<td>Mixed-sex dyad, female employer</td>
</tr>
<tr>
<td>No stereotype threat</td>
</tr>
<tr>
<td>Same-sex male dyads</td>
</tr>
<tr>
<td>Same-sex female dyads</td>
</tr>
<tr>
<td>Mixed-sex dyad, male employer</td>
</tr>
<tr>
<td>Mixed-sex dyad, female employer</td>
</tr>
</tbody>
</table>

Note: In each of these dyads, one person was the employer and the other was the employee.
might be due to Type II errors. For example, it is still possible that power or mode of communication qualifies the previously described effects of stereotype threat in mixed-sex pairs, but those interactions were not significant in this experiment. Although the researchers had originally conducted a power analysis and had planned to collect data from 300 participants, recruiting that many participants proved to be a complex challenge due to the requirements of the study and the size of the participant pool. However, the sample size of this study was acceptable to test the five main effects (Horton & Lipsitz, 1999), of which only one, power, was significant. Future research with larger samples would be helpful to test the significance of the higher order interactions and attempt to replicate the results which were significant in this study.

It is also possible that the lack of significant results regarding the mode of communication was due to the ways in which the e-mail negotiations were conducted in this particular experiment. Because participants in the e-mail condition met in person before they were sent to separate rooms to e-mail each other, these participants were aware of their partner’s sex. If participants in the e-mail condition had been entirely unaware of their partner’s sex, then the salience of sex might have decreased and the negotiation might have been an especially “weak” situation with regard to prescriptive stereotypes. However, in real world e-mail communications, writers are rarely anonymous, their sex can often be ascertained quite easily from their e-mail address, and employers usually know the sex of potential employees during the negotiation process. Although the brief face-to-face meeting before the e-mail negotiation might not have decreased external validity that much, the 15-min time limit for the negotiations most likely did. Although most pairs were able to reach an agreement within this time limit, six of the pairs in the e-mail condition and only one pair in the face-to-face condition failed to reach an agreement. The difficulty in the e-mail condition was most likely due to the nature of e-mail; it takes longer for e-mails to be typed, sent, and received than for people to speak to one another. The external validity of the experiment could have been improved by allowing more time for the negotiations to take place, particularly the e-mail negotiations.

The structure of the salary goal sheets might also have weakened the external validity of the experiment. Although participants did not see their partner’s goal sheet, they might have realized that agreeing on a specific salary would guarantee that they would split the financial reward evenly ($5 each). Although only 18 out of 48 pairs agreed to a $5 split, it was enough to bring the salary goal sheet into question. In true salary negotiations, there is no prestated salary that can be objectively defined as equally fair to the employer and the employee. Instead, there is a certain amount of subjectivity as to what really is fair to both parties. If the participants described their goal sheets to each other, some might have decided to agree upon the objectively fair outcome rather than negotiating more aggressively under ambiguous circumstances. However, the videotapes and e-mail transcripts of the negotiations revealed that in only two cases did the pairs discuss their salary goal sheets and agree to a $5 split rather than negotiating with each other.

Although some lab experiments have studied the negotiation outcome of just one participant at a time because the other person was a confederate following a negotiation script (Barron, 2003), this experiment studied both people during the negotiation to increase external validity as other researchers have done (Kray et al., 2001). However, this methodological choice made it difficult to interpret the cause of the stereotype threat effects in the mixed-sex dyads. Because these negotiations were zero-sum interactions, one partner doing better by a certain amount meant that their partner must have done worse by exactly the same amount. Therefore, it was difficult to know in mixed-sex dyads if stereotype threat caused the men to negotiate more successfully, the women to negotiate less successfully, or both. For example, did men show

---

**FIGURE 1**

**The Effects of Stereotype Threat on the Monetary Outcome of Men and Women in Mixed and Same-Sex Pairs**

![Graph showing the effects of stereotype threat on monetary outcomes for mixed and same-sex pairs.](image-url)

Note: The bars are upside down to illustrate how much worse than the goal the outcome was. The asterisks indicate that the two bars were significantly different from each other.
Stereotype Threat and Negotiation | McCormick and Morris

the positive effects of stereotype boost because they negotiated more aggressively or did they simply benefit from negotiating with women who negotiated poorly due to stereotype threat? Although the current study could not distinguish between these two explanations, it seems likely that both explanations are plausible. Past research with confederates has shown evidence of stereotype threat negatively impacting women’s negotiation outcomes (Tellhed & Björkland, 2011) and positively impacting men’s (Kray et al., 2001). However, regardless of which person’s behavior drove the effect in the current experiment, the outcome was the same; men ended up with better negotiation outcomes and women ended up with worse negotiation outcomes when stereotype threat was activated in a mixed-sex pair.

Although the focus of the present study was on stereotype threat during salary negotiations, many other factors are responsible for the sex difference in salaries. According to Blau and Kahn (2007), almost 60% of the wage gap in the United States can be explained by known factors related to occupational choices. For example, women tend to be employed in lower paid professions than men. Women are also more likely to work part-time than men and they work fewer years than men due to childcare and maternity leave. However, over 40% of the wage gap cannot be accounted for in a measurable way. These immeasurable explanations likely include sex discrimination and sex differences in salary negotiations (Blau & Kahn, 2007; Stuhlmacher & Walters, 1999).

An additional limitation of the present study is that the results may be culture-specific. The process of negotiation varies cross-culturally (Adair & Brett, 2005). In Western cultures, negotiators typically use direct communication and their meaning is explicitly stated. Whereas in Eastern cultures, negotiators tend to use indirect communication characterized by implicit embedded meaning that is not clearly stated (Gibson, 1998). It is likely that these different communication norms affect the process of salary negotiations for both men and women. Because the present study relied on a Western sample of participants, these results should not be generalized to non-Western cultures.

Implications and Future Directions

The present research suggested that women under-perform in mixed-sex negotiations when the negotiation task is described as being diagnostic of their abilities, a description which activates stereotype threat. Although this might lead to the assumption that women are always at a disadvantage in mixed-sex negotiations, because negotiations are probably assumed to be a reflection of skill, there are situations in which women’s outcomes are not impaired by sex stereotypes. Describing the negotiation task as challenging and diagnostic is an implicit way to activate stereotype threat whereas telling participants that men are better negotiators than women is an explicit way to activate stereotype threat. Although the implicit stereotype threat manipulation used in the present study caused men to outperform women, when stereotype threat is explicitly activated, women display reactance and actually outperformed men (Kray et al., 2001), especially when they do not believe that ability to negotiate is an innate characteristic (Kray & Haselhuhn, 2007). Furthermore, if people are reminded of the stereotypically feminine characteristics more conducive to effective negotiation such as listening and communication skills, women negotiate more effectively than men do (Kray et al., 2002). Taking the results of the current study and past research into consideration, one could argue that women would be more effective negotiators in the real world if they are either reminded of the feminine qualities that are needed for negotiation or they are so conscious of the negative stereotypes that they react against them. When stereotype threat is subtly activated in mixed-sex pairs, women are more likely to underperform.

It is possible that the effects of stereotype threat in mixed-sex pairs was related to the fact that the hypothetical job was in a science, technology, engineering, or mathematics (STEM) field, and women are underrepresented in that field.
(U.S. Department of Commerce, 2011). It has been argued that stereotype threat is ever-present for people who are in an environment where they are in the minority such as women in STEM fields (Holleran, Whitehead, Shmader, & Mehl, 2011). In fact, just mentioning a STEM-related job in this experiment might have, in and of itself, activated stereotype threat for women and stereotype boost for men. Thus, the stereotype threat manipulation used in the present experiment might have actually removed the STEM-related threat rather than adding a threat. As Figure 1 illustrates, the removal of the threat impacted the results; when in mixed-sex pairs, women did better and men did worse when the task was described as easy for novices and not an assessment of their skills. Future research could explore whether the effects of stereotype threat change when the salary negotiation is for a job that is more stereotypically feminine.

Although the current experiment did find significant effects of power on negotiation outcomes and expectations, power was only manipulated in a hypothetical sense. Participants were told to imagine that they were employers or employees, but the participant in the employer role had no true power over the participant in the employee role during their interaction. In real salary negotiations, the employer has real power in that he or she can put an offer on the table and say “take it or leave it.” The employee may not negotiate as aggressively for fear of making a bad impression on a future boss, and the employee may be more desperate for the job than the employer is to fill the position. In this experiment, hypothetical power was manipulated, but actual power was not; the two participants needed to come to an agreement in order to earn some portion of the $10, and the employer could not simply decide to keep the $10 for him or herself. Nevertheless, despite the hypothetical manipulation of power, participants in the more powerful role must have felt somewhat more powerful because they expected to be more successful and actually were more successful than participants in the less powerful role. In a future experiment in which actual power is manipulated, it is likely that these effects would be stronger.

The lack of a difference between e-mail and face-to-face negotiation outcomes has interesting implications for negotiations in the real world. As virtual forms of communication have become increasingly prevalent, more negotiations are likely to occur via e-mail. Employees and employers may wonder if their outcomes are affected by the mode of communication they choose, and this research suggested that this might not be the case. E-mail also may not minimize the effects of stereotype threat. However, future research, with larger sample sizes, should examine whether e-mail negotiations decrease the effects of stereotype threat and sex differences when participants never meet face-to-face and sex is less salient or even when the sex of the other participant is entirely unknown.

Conclusion

The present research showed that stereotype threat improves men’s outcomes and impairs women’s outcomes in mixed-sex negotiations. In support of past research, power was a strong predictor of how successfully individuals negotiated. Although women negotiated more effectively when they were in a position of power, having power did not make them immune to the negative impact of stereotype threat when they negotiated with men. Future research, with larger sample sizes, should further test whether power and mode of communication, either face-to-face or e-mail, alters the effects of stereotype threat on men and women. As more communication takes place virtually, it will become increasingly important to explore these questions, and the answers may shed light on ways to decrease the sex salary gap.

References

Burgess, D., & Borgida, E. (1999). Who women are, who women should be:
Stereotype Threat and Negotiation


Author Note: Jasmine McCormick and Wendy L. Morris, Department of Psychology, McDaniel College, MD.

Funding for this project was made possible through a Psi Chi Faculty Advisor Research Grant awarded to Wendy Morris. We would like to thank our research assistant, Katya Spitznagel, and also Laura Sherman who provided feedback on the data analysis and an earlier version of this article.

Correspondence concerning this article should be sent to Wendy L. Morris, Department of Psychology, McDaniel College, 2 College Hall, Westminster, MD 21157. E-mail: wmorris@mcdaniel.edu
Do We Have Fun When Time Flies?
Matthew Pageau, Saint Joseph’s University
Seth Surgan*, Worcester State University

**ABSTRACT.** The purpose of the present study was to assess the effects of distortions in time perception and the complexity of a task on mood and enjoyment of the task. Further, researchers sought to study factors that may affect time perception (e.g., immersion, task complexity, objective task duration). Seventy-seven undergraduate students participated in 1 of 6 conditions within a 2 (complexity) x 3 (duration) design. Although the experimental manipulations did not create the anticipated effects, subsequent analyses supported the hypotheses that, when time was perceived as passing quickly, participants experienced more positive mood change, \( t(75) = 3.92, p < .001, \eta^2 = .17 \). Further, immersion was positively correlated with distortions in time perception, \( r(75) = .40, p < .001, r^2 = .16 \), and mood change, \( r(75) = .23, p = .02, r^2 = .05 \). Regression analysis revealed that distortions in time perception mediated the relationship between immersion and mood change. Theoretical implications for self-perception theory and the concept of flow are discussed.

Time is a basic feature of human experience. Across societies, people have internalized different models of time through which they connect events, organize their lives, and attempt to come to terms with fundamental questions about life, death, and existence (Levine, 1997). As a species, people have many models of time that they use in everyday life. These models may vary according to the direction in which the *arrow of time* points (Núñez & Sweetser, 2006) or even whether time is seen as flowing in a linear way at all (Lee, 1950; Whorf, 1956).

**Objective and Subjective Time**
Despite this variety across societies, individuals’ social and economic lives are structured in very real ways by both their efforts to standardize time (Levine, 1997; Sorokin, 1943) and their expectation that others structure time similarly. For example, when a friend promises to meet in an hour, it is not expected to have to wonder if this will be a fast hour or a slow hour. However, when someone arrives, one might (if the situation calls for it) judge the friend to be early, on time, or late based upon this objective standard of time passing. This is also true for shipments, transportation schedules, media events, communication over large distances, and the many varieties of coordination that allow the modern world to function. During interactions with other people, the expectation is that individuals will operate according to standardized objective time, in which the duration of every second is the same.

However, such standards and expectations may not apply in an individual’s own psychological sphere. In a person’s everyday experience, moments stretch or shrink, and may not be amenable to division into elementary units such as minutes or seconds. When people reflect on their experiences and try to estimate how much time has passed, it is easy to see the potential mismatch between their subjective experience of time and the standardized version. For example, someone immersed in a good

---

*Faculty mentor
Do We Have Fun When Time Flies?

who experienced time passing normally or more reported enjoying a task more than participants to feel time passing more quickly than normal demonstrated that participants who were made perception all seem to be related. Sackett, Emotions, enjoyment, and distortions in time (Ratcliffe, 2012).

We Also Have Fun When Time Flies

Everyday discourse frequently reflects commonsense beliefs. Everyone seems to know that “time flies when you are having fun.” Recent research has demonstrated that this is more than just a saying. For a variety of reasons, enjoyable experiences are felt to pass more quickly than unpleasant (or less pleasant) ones (Droit-Volet & Meck, 2007; Gable & Poole; 2012; Iwamoto & Hoshiyama, 2011). For example, Iwamoto and Hoshiyama (2011) found that participants who reported more interest in completing a 24-piece jigsaw puzzle also felt the puzzle-solving task to pass more quickly than did participants who reported little interest in the task. Similarly, Gable and Poole (2012) found that time seemed to pass more quickly for participants who were pursuing intrinsically rewarding activities (e.g., family time, exercise, or eating a favorite food).

Likewise, other research has demonstrated that time seems to drag when individuals are not enjoying themselves (Danckert & Allman, 2005; O’Brien, Anastasio, & Bushman, 2011; Ratcliffe, 2012). Danckert and Allman (2005) found that participants who were bored experienced time as passing more slowly than those who were not bored. Similarly, O’Brien et al. (2011) found that participants who ascribed greater value to their personal time overestimated the length of a dull task, reported less interest in it, and walked away from the laboratory more quickly once the dull task was completed. In instances of depression, time is often described as slowing down or stopping (Ratcliffe, 2012).

We Also Have Fun When Time Flies

Emotions, enjoyment, and distortions in time perception all seem to be related. Sackett, Meyvis, Nelson, Converse, & Sackett (2010) demonstrated that participants who were made to feel time passing more quickly than normal reported enjoying a task more than participants who experienced time passing normally or more slowly than normal (Sackett et al., 2010). That is, not only does time seem to fly when one is having fun, but research has suggested that people also have fun when time flies. Sackett et al. (2010) proposed that time perception is used as a metacognitive cue where subjective temporal experience is referenced to establish hedonic evaluations of experiences. In other words, a person’s awareness of time having passed either more quickly or more slowly than usual is a factor that is taken into account in the process of formulating judgments regarding how much or little they enjoyed an experience.

This finding was in line with both the James-Lange theory of emotions (James, 1884) and self-perception theory (SPT; Bem, 1972), which states that people may come to know their own emotional states by inferring them from observation of their own behavior. Over the course of decades, James Laird (Duclos & Laird, 2001; Laird, 1974; Laird & Bresler, 1992; Schnall, Abrahamson, & Laird, 2002) has repeatedly demonstrated that emotional experiences are contingent upon a subconscious awareness of expressive behaviors. For example, smiling or furrowing the brow seems to influence the degree to which an individual feels happy. Instead of the commonsense belief that people are smiling because they are happy, SPT suggests that the inference is more along the lines of “I am smiling, so I must be experiencing happiness” (Laird, 1974). Sackett et al. (2010) demonstrated that a subconscious awareness of distortions in time perception serves a similar purpose in the construction of emotional experiences. That is, “Time flew, so I must have been enjoying myself.”

The Importance of the Task

Researchers seeking to understand the fluidity of time perception have focused on cognitive factors that are often linked to the nature of the activities that occupy people’s time (e.g., task complexity and other demands on cognitive resources; Iwamoto & Hoshiyama, 2011; Ornstein, 1969; Pariyadath & Eagleman, 2007). This body of research has indicated that time tends to drag as cognitive demands increase. This idea has been summarized by Ornstein’s (1969) storage-size metaphor, which claims that subjective duration is represented through metaphorical space that information takes up when encoded, with more complex information requiring more metaphorical space. Ornstein conducted five experiments that supported his theory that cognitive demands are linked to temporal
perception and that subjective time seems to slow down as cognitive demands increase.

In an applied setting, Stetson, Fiesta, and Eagleman (2007) demonstrated that people in a life-threatening situation remember the event taking longer than it really did. The authors suggested that this dilation of subjective duration is due to increased encoding of information during the event. Along similar lines, Vohs and Schmeichel (2003) demonstrated that demands for self-regulation increase the perceived duration of events. These researchers asked some of their participants to suppress their reactions during exposure to emotional stimuli. Other participants were not asked to self-regulate. In their study, self-regulators estimated that the task took significantly longer than nonregulators. Finally, Iwamoto and Hoshijama (2011) demonstrated that task complexity is an important factor in time perception. All other things being equal, more complex tasks are experienced as taking longer than less complex tasks. These findings all converged on the notion that perception of time fluctuates with the level of cognitive demand.

Getting “Into It”: Immersion, Flow, and the Passage of Time
Other research has suggested that objective characteristics of the task are not sufficient to explain distortions in time perception. This work focuses on the interaction between the person and the task at hand. Task immersion is one factor that is strongly associated with distortions in time perception. Immersion is a feeling of full concentration including effortless attention and focus on the task at hand, emotional involvement, and blocking out of external distracters.

Nakamora and Csikszentmihalyi (2001) suggested that task immersion is most likely to happen when there is a good balance between the difficulty of the task and the skill level of the individual performing the task. In this situation (entailing demand-skill correspondence and task immersion), the experience of flow, an intrinsically rewarding state of intense concentration on the present task, is likely to occur (Csikszentmihalyi, 1975). Flow is an outcome from the act of being immersed into a task where one receives an intrinsic reward from the experience. Flow experiences include immersion, but not every experience of immersion is a flow experience. One especially relevant characteristic of flow experience is the tendency for time to pass more quickly than normal. That is, time often seems to fly when a person is immersed in a task that optimally matches their skill level.

The Present Study
The current study tested several hypotheses, which replicated and synthesized previous research. First, we hypothesized that participants who experienced an increase in the perceived speed of time (i.e., time flying) while performing a reading task would demonstrate an increase in positive subjective experience (i.e., mood and task enjoyment). We suggested that this would be because of the dynamics of self-perception. The feeling that time has passed more quickly (or more slowly) than usual would be used as a metacognitive cue within the process of formulating an experience as suggested by the James-Lange theory of emotion and Self-Percetion Theory and demonstrated by Sackett et al. (2010). Next, we hypothesized that participants who performed a more complex task would experience significantly less positive mood and less enjoyment than participants performing a less complex task. We suggested that this would be because participants would find it more difficult to become immersed in more complex tasks. Also, as Iwamoto and Hoshijama (2011) demonstrated, increased complexity tends to make time drag, and lower complexity makes time pass quickly. Finally, we hypothesized an interaction between changes in time perception and task complexity. We suggested that this interaction would exist because, consistent with Csikszentmihalyi’s work on flow, the lack of immersion in a more complex task would create a dilation of subjective time (i.e., the feeling that time has dragged) that would counteract our attempt to induce the feeling that time has flown. On the other hand, greater immersion in a less complex task was expected to enhance the effect of the feeling that time has flown, producing an even stronger increase in positive subjective experience.

Method
Participants
Seventy-seven undergraduate students from an eastern university were recruited through the online sign-up SONA software system (55 women, 22 men; $M_{age} = 20.60, SD = 6.84$). Most of our participants (63 of 77) self-identified as European American. Only five participants self-identified as either American/Alaskan Native, Asian, or African American. The remaining nine participants self-identified as other. Participants were compensated with credit toward fulfillment of the research
Do We Have Fun When Time Flies?

Pageau and Surgan

requirement for their general psychology course (if applicable) or extra credit in another class (if offered).

Research Design
The present research used a 2 x 3 between-subjects design in which all participants agreed to complete a 10-min reading task. The independent variables were task complexity (low, high) and distortions in time perception (flying, normal, dragging). We manipulated the complexity of the reading task by asking participants in the low complexity condition to read a passage normally. Participants in the high complexity condition read the same passage with the text presented in mirror image. We also sought to manipulate our participants’ experience of time. We operationalized this in terms of task duration. Although all participants believed that the reading task would last 10 min, we manipulated the actual duration of the task (5, 10, or 15 min) to create the feeling that time had flown in the 5-min condition or dragged in the 15-min condition. A similar procedure successfully produced distortions in time perception previously (cf. Sackett et al., 2010).

Task enjoyment and mood were the dependent variables of interest. To check whether we had successfully manipulated participants’ time perception and help support our theoretical explanation for the hypothesized effects of the independent variables, we assessed participants’ immersion in the experimental task and the degree to which participants felt time either flying or dragging.

Materials
Brief Mood Introspection Scale. (BMIS; Mayer & Gaschke, 1988) A 10-item, 7-point Likert-type scale was used to assess mood before and after the reading passage. Scores range from 10 to 70, with higher scores representing more positive mood. The BMIS has a Cronbach’s alpha of .80. For the current sample, Cronbach’s alpha was .80 on the pretest and .85 on the posttest.

Reading passage. The reading passage included Chapters 5 and 6 from Todd Strasser’s (1981) novel The Wave. It was printed in both normal (low complexity) and mirror image (high complexity) format. We chose to use a reading task for two reasons. First, we chose an excerpt from a novel in the hope that this would facilitate task immersion. Our second reason for choosing a reading task was that it lent itself to manipulations of complexity, through necessary mental flipping to understand the words, similar to that used in other current research in cognitive psychology (Koriat, Norman, & Kimchi, 1991; Zhang, Qui, Huang, Zhang, & Bao, 2009). Presenting the text in mirror image is a straightforward way of making the reading task more complex and one that we thought would highlight the presumed connection between task complexity and immersion.

Overall Task Enjoyment Scale (OTES). A study-specific scale was created to measure enjoyment of the task and distortions of time perception. The first item measured enjoyment over a 9-point Likert-type scale, with higher scores indicating greater enjoyment. Distortions in time perception, the second item, were assessed over a 9-point Likert-type scale, with higher scores indicating time being perceived as going by more quickly than usual. After data collection was complete, scores on this scale were centered around 0 so that negative scores indicated time passing more slowly than usual and positive scores indicated time passing more quickly.

Task Immersion Scale (TIS). Immersion was measured using a modified version of the Immersive Tendency Questionnaire (ITQ), created by Witmer and Singer (1998). The original ITQ consisted of 19 items, each answered using a 5-point Likert-type scale, with a Cronbach’s alpha of .78. Weibel, Wissmath, and Mast (2010) found the ITQ to include two factors: emotional involvement and absorption. The present study focused on absorption. Therefore we used a set of four items based on the absorption subscale proposed by Weibel et al. Items included “While performing the reading task, to what extent did you feel you lost track of things around you?” and “How effectively were you able to block out external factors while you were reading?” Each of the four items was answered using a 5-point Likert-type scale, with higher total scores representing greater task immersion. The first item measured enjoyment over a 9-point Likert-type scale, with higher scores indicating time passing more quickly.

Procedure
After human subjects review board approval (ID#1213-17-01) was obtained, researchers constructed an environment where time manipulation would not be confounded by other variables (e.g., temperature, time awareness). To do so, a room was set up without clocks, and at a stable temperature. After providing informed consent, a group of participants was instructed to fill out a demographic scale and the BMIS to acquire a baseline mood score. The examiner then explained to the
participants that they would be asked to read a passage to the best of their ability for 10 min. In reality, participants were assigned to reading the passage for either 5, 10, or 15 min. Following completion of the reading task, participants were instructed to complete a second BMIS, the OTES, and the TIS. Participants received a debriefing sheet to emphasize that their mood might have changed as a result of an experimenter-implemented time manipulation, but the normal rate of time had not changed.

Results
Was Time Perception Effectively Manipulated?
In the present study, the duration and complexity of the reading task were manipulated for the purpose of producing distortions in time perception. It was hypothesized that distortions in time perception would affect both participants’ mood and their enjoyment of the reading task. Before examining the effects of the experimental manipulations on mood and enjoyment, it was important to assess the effectiveness of our experimental manipulations on time perception. To do this, we performed a factorial Analysis of Variance (ANOVA) using scores on the second item of the OTES. This item asked participants to reflect back on the 10 minutes they spent performing the reading task and indicate whether that time seemed to pass more slowly or quickly than normal. Significant main effects on time perception were observed for task duration, $F(2, 71) = 10.29, p < .001, \eta^2 = .19$, and task complexity, $F(1, 71) = 4.12, p = .05, \eta^2 = .04$. As expected, time appeared to pass the fastest for participants in the 5-min group (regardless of complexity) and slowest for those in the 15-min group. Similarly, time passed significantly more quickly for participants in the low complexity condition, regardless of task duration.

The analyses also revealed a significant interaction between task complexity and task duration on time perception, $F(2, 71) = 5.60, p = .002, \eta^2 = .10$. Although time moved faster than normal for participants in the 5-min condition (regardless of complexity) and slower than normal for participants in the 15-min condition (regardless of complexity), participants in the 10-min condition were split. Among this group, participants in the low complexity condition experienced time flying, while time dragged for participants in the high complexity group. Interestingly, time seemed to have passed most quickly for those in the 5-min high complexity condition, and second fastest for those in the 10-min low complexity condition (see Figures 1 and 2).

A Closer Look: What Happens When Time Flies?
Despite our apparent success at manipulating time perception by changing the duration of the task, we were puzzled by the lack of any consequent effect of duration on either mood, $F(2, 71) = 0.59, p = .56, \eta^2 = .01$, or enjoyment, $F(2, 71) = 1.51, p = .23, \eta^2 = .03$. This contradicted previous research (Sackett et al., 2010), which indicated the hedonic value of distortions in time perception.

A promising finding regarding the effect of
distortions in time perception on mood emerged when we divided our participants into two groups using OTES scores: those who experienced time passing more quickly than usual and those who did not, regardless of the duration and complexity of the task being performed. We categorized participants as having experienced time passing more quickly if they produced a positive score on the OTES (i.e., the fast group). Those who did not produce a positive score on the OTES were categorized as having experienced time passing more slowly than usual (i.e., the slow group). We ran an independent-samples $t$ test to compare mood change scores from the fast group ($n = 33, M = 1.24, SD = 7.67$) and slow group ($n = 44, M = -5.45, SD = 7.24$). The results indicated that participants who experienced time flying had significantly more positive changes in mood scores than those who did not, $t(75) = 3.92, p < .001$, $\eta^2 = .17$, thus providing support for the hypothesized effect of distortions in time perception on mood. In addition, we observed the hypothesized main effect of task complexity on mood, $F(1,71) = 14.90, p < .001$, $\eta^2 = .16$. Participants in the high complexity group ($M = -5.69, SD = 8.18$) experienced significantly less positive mood change than participants in the low complexity group ($M = 0.61, SD = 6.73$).

We observed a similar pattern for enjoyment, with participants in the fast group ($M = 4.76, SD = 2.05$) showing slightly higher scores than participants in the slow group ($M = 4.14, SD = 2.54$). However, this difference was not statistically significant, $t(75) = 1.15, p = .22, \eta^2 = .02$. It seems that task complexity alone accounted for the degree to which participants enjoyed their assigned reading task, $F(2,71) = 20.73, p < .001$, $\eta^2 = .22$.

Although there was no significant main effect of task duration on mood change or a significant duration x complexity interaction, we were able to detect significant effects when we split our participants according to their experienced distortions in time perception. By approaching the data through this lens, we were able to detect a significant difference in mood change scores between participants who experienced time flying and those who did not. Specifically, participants who experienced time flying demonstrated more positive mood change than participants who did not. These results indicated that distortions in time perception seemed to be having their hypothesized effects on mood. However, that effect could not be traced back to our manipulation of task duration. This suggests that the time manipulation might not have been as successful as our initial manipulation suggested. This led us to look even more closely at whether manipulating task duration was an effective way of inducing distortions in time perception.

Did Task Duration Reliably Affect Time Perception?

In our experimental design, we created three task duration conditions: 5, 10, and 15 min. The intention was that the 10-min condition would be a control condition and that participants in that group, on average, would report no distortions in time perception. A one-sample $t$ test supported our assumption by demonstrating that posttest OTES scores for this group of participants ($M = -0.24, SD = 2.55$) were not significantly different from 0, $t(24) = -0.47, p = .64, d = -.09$. Another indication that this group worked well as a control group was the fact that 12 out of 25 participants (48%) produced positive OTES scores, and 13 out of 25 (52%) did not. A $\chi^2$ goodness of fit test showed that this distribution was not significantly different than what was expected by chance, $\chi^2(1, N = 25) = .04, p = .84$. This even split indicated that participants in this condition did not reliably experience any specific kind of distortion in their perception of time. The split also closely mirrored the number of participants in the 10-min, low complexity group (11) and 10-min, high complexity group (14), indicating that the 10-min condition was a ‘clean slate’ upon which the effects of task complexity could come through most clearly.

Participants in the 15-min condition behaved very differently. By extending the duration of the reading task beyond the expected 10 min, we hoped to induce the feeling that time had passed more slowly than usual. This would manifest itself as negative OTES scores. A one-sample $t$ test supported this expectation by demonstrating that OTES scores for this group ($M = -1.13, SD = 1.87$) were significantly below 0, $t(23) = -2.94, p < .001, d = -.60$. Only 3 out of 24 (13%) participants in this group produced positive OTES scores, with another four participants reporting that time passed normally. Seventeen out of 24 participants in this group (70.83%) produced negative OTES scores. This split was consistent across complexity levels. Nine out of 13 participants in the low complexity condition and 8 out of 11 in the high complexity condition produced negative OTES scores. A $\chi^2$ goodness of fit test showed that this distribution was significantly different than what would be expected by chance, $\chi^2(1, N = 24) = 4.17, p = .04$. 

Pageau and Surgan
These findings suggested that participants in this condition reliably experienced time passing more slowly than normal, regardless of task complexity.

However, participants in the 5-min condition did not follow our expectations as clearly. We expected these participants to experience time passing more quickly than normal, as shown by positive OTES scores. A one-sample t test supported this expectation. Scores for this group (M = 1.36, SD = 2.02) were significantly above 0, t(27) = 3.55, p < .001, d = .67. This suggested that participants in this condition not only experienced the kind of distortions in time perception that we expected, but that the effect was even stronger than in the 15-min condition. Looking more closely at the data, however, only 18 out of 28 (64%) of participants in the 5-min group produced positive OTES scores. More than one third of participants in this group did not produce positive OTES scores. This split was consistent across complexity conditions, with 9 out of 14 participants in each complexity condition producing positive OTES scores. A χ² goodness of fit showed that this distribution was not significantly different than what would be expected by chance, χ²(1, N = 28) = 2.29, p = .13. This suggested that, regardless of task complexity, participants in this condition did not reliably experience the expected distortions (i.e., shortening) of time perception.

Based on this participant-level manipulation check, we believe that we did not successfully induce distortions in time perception by manipulating task duration. Specifically, participants in the 5-min condition did not experience time flying at a rate that was significantly greater than chance. This accounts for why we did not see a significant main effect for task duration on mood and why we did observe the hypothesized significant difference in mood change when participants were grouped based on their OTES scores. This, together with the fact that there were no significant effects of either duration, F(2, 71) = 0.63, p = .53, η² = .02, or complexity, F(1, 71) = .25, p = .62, η² < .01, on immersion scores, and no significant interaction, F(2, 71) = .38, p = .68, η² = .01 (see Figure 3), led us to set our experimental manipulations aside and analyze the functional relationships between time perception, immersion, and mood change more directly.

Building a Functional Model

Although it was difficult to see how distortions in time perception, immersion, and mood change fit together when looking through the lens of our manipulations of task duration and complexity, a clearer picture emerged when we analyzed the direct measures of those variables. Increasing task immersion, as reflected in TIS scores, was associated with the feeling of time passing more quickly, r(75) = .40, p < .001, r² = .16, and increasing positive changes in mood, r(75) = .25, p = .02, r² = .05. Following up on these preliminary findings, we performed regression analyses to determine whether distortions in time perception and task immersion were significant predictors of mood change.

Using stepwise regression analyses, we determined that immersion was a significant predictor of mood change, β = .23, t(76) = 2.05, p = .04. Greater task immersion predicted positive changes in mood. The power of task immersion to predict mood change was weakened considerably when distortions in time perception were included in the model, β = .06, t(76) = .51, p = .61. In this context, immersion alone was no longer a significant predictor of mood change. Subsequent mediation analysis performed using Baron and Kenny’s (1986) four-step technique revealed that distortions in time perception mediated the relationship between immersion and mood change (see Figure 3 and Table 1). That is, immersion positively predicted distortions in time perception, β = .40, t(76) = 3.79, p < .001, and distortions in time perception predicted mood change, while controlling for the effect of immersion, β = .43, t(76) = 3.80, p < .001. See Figure 4 for an illustration of these functional relationships.

Discussion

Previous research has established that manipulations in task complexity distort time perception...
Do We Have Fun When Time Flies?

(Iwamoto & Hoshiyama, 2011; Ornstein, 1969; Pariyadath & Eagleman, 2007; Stetson et al., 2007; Vohs & Schmeichel, 2003). In addition, manipulating task duration has been demonstrated to produce distortions in time perception, which can affect participants’ enjoyment of the experimental task (Sackett et al., 2010). The present study synthesized these works by examining the effects of both task duration and task complexity on task enjoyment and mood. It was hypothesized that duration and complexity would affect mood and enjoyment by producing distortions in time perception. Following self-perception theory (Bem, 1972; Laird, 1974), we expected participants to perceive and interpret distortions in time perceptions as signs of their own emotional state and task enjoyment.

Initial findings indicated that task complexity affected time perception, task enjoyment, and participants’ mood in the expected directions. In the mirror-image condition, participants experienced time as passing more slowly, reported enjoying the task significantly less (see Figure 5), and experienced significantly less positive mood change (see Figure 6) than participants in the low complexity condition. These findings directly supported our hypotheses and corresponded with previous research findings.

Initial findings regarding task duration were not as clear-cut. As expected, the duration of the reading task affected participants’ perception of time but did not affect either their mood or enjoyment of the reading task. This mixed set of findings prompted us to perform a series of manipulation checks, which led us to abandon the factorial structure of the original experiment and analyze direct measures of the relevant variables (i.e., distortions in time perception, immersion, mood change, and task enjoyment).

Subsequent analyses demonstrated that people do, in fact, have fun when time flies. The experience of time flying was found to predict positive changes in mood. Further, the present study has also demonstrated that distortions in time perception are predicted by task immersion. When a person becomes immersed in a task, they are likely to experience time as passing more quickly than usual, which can produce positive mood changes. In other words, becoming immersed in a task sets the stage for improved mood, but the experience of time passing quickly is a necessary ingredient. Without distortions in time perception, simply becoming immersed in a task is not sufficient to produce better mood. In this sense, our data suggested that people do tend to enjoy themselves more when time seems to fly. This directly supported our primary hypothesis.

The important role of distortions in time perception is consistent with the theoretical basis of the present study, the James-Lange theory of emotion (James, 1884) and self-perception theory (SPT; Bem, 1972; Laird, 1974). According to this approach, the awareness of time having passed more quickly than usual is taken as part of the experience of improved mood. Just as Laird (1974) suggested that a person may infer their own happiness from an awareness that they are smiling, our participants seemed to experience enhanced mood because of their awareness that they felt time flying. This corroborates the empirical support for SPT supplied by Laird and his colleagues (Duclos & Laird, 2001; Laird & Bresler, 1992; Schnall et al., 2002) as well as by Sackett et al., 2010, and

### TABLE 1

<table>
<thead>
<tr>
<th>Predictors</th>
<th>B</th>
<th>95% CI</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immersion</td>
<td>.23*</td>
<td>[0.02, 1.15]</td>
<td>12.53</td>
<td>3.20</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immersion</td>
<td>.06</td>
<td>[0.42, -0.72]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distortions in time perception</td>
<td>.43**</td>
<td>[0.07, 2.22]</td>
<td>0.65</td>
<td>2.38</td>
</tr>
<tr>
<td><strong>Outcome—Mood Change</strong></td>
<td>-2.58</td>
<td>8.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $R^2 = $Proportion of outcome variable variance explained by predictors. $p < .05$, $^*p < .01$, Cohen’s $f$ = effect size.

Step 1 to Step 2: $R^2 = .05, (F, 1/75) = 4.18, p = .04, Cohen’s f = .06$.
Step 2: $R^2 = .16, (F, 1/74) = 14.45, Cohen’s f = .18$.
Step 1 to Step $2: R^2 = .21, (F, 2/74) = 9.69, p < .001, Cohen’s f = .27$.

### FIGURE 4

Functional Relationships Between Immersion, Distortions in Time Perception, and Mood Change

![Diagram](image)

Note. Distortions in time perception during the reading task mediated the relationship between immersion and mood change. On its own, immersion (positively) predicted mood change ($p = .04$). When distortion in time perception was added to the model, this relationship disappeared ($p = .51$). The significant predictive relationships between immersion and distortion in time perception ($p < .001$), as well as between distortion in time perception and mood change ($p < .001$), demonstrated the mediational role of distortion in time perception. Positive scores represent positive change in mood. $N = 77$. 
suggests that time perception may be an integral but relatively unexplored precursor of positive mood change.

Although we were able to build a functional model of mood change that supported our hypotheses and corroborated SPT, the findings regarding task enjoyment did not support our hypotheses. Enjoyment was not affected by task duration and was not predicted by distortions in time perception. We were unable to integrate enjoyment into our functional model. Task complexity was the only factor that influenced how much participants enjoyed the reading task. Both task enjoyment and mood were uniformly low for the high complexity group, regardless of duration. There are several potential methodological explanations for this finding. Our complexity manipulation (i.e., presenting the stimulus reading in mirror-image) might have been too demanding, given the participants’ level of investment. The novelty of the task (both compared to what participants do in their everyday lives and as participants in other studies) might also have led the participants to simply react to the task, rather than engage with it. A more subtle manipulation (i.e., a task that does not differ from the low complexity version or from everyday activities in such an obvious way) might be more suitable for future research. Finally, another reason for this negative result might be a lack of power in the experiment. Practical constraints (including the looming graduation of the first author and a lack of research assistants to help administer the study) limited the sample size and, consequently, the power of the analyses.

From Mood to Flow

Although the potential mismatch between the demands of our experimental task and the interests and motivations of our participants might have undermined some of our initial hypotheses regarding the effect of task complexity, this also highlighted the potential relevance of the concept of flow (Csikszentmihalyi, 1975). Flow involves full engagement in a task and is characterized by both positive affect and a feeling that time has passed quickly. Our results suggested a central role of time perception in the experience of flow and the intrinsic emotional rewards that are characteristic of such experiences. Specifically, the positive subjective experience associated with flow seems to depend upon the distortion of time perception. In both Csikszentmihalyi’s theoretical elaboration of flow and in our findings, positive changes in mood are predicted by task immersion. The present study contributed to the understanding of the process by which flow experiences emerge by demonstrating that distortions in time perception mediate the predictive relationship between task immersion and positive mood change. In this way, the same process of registering changes in a person’s own subjective flow of time that may lie at the center of mood change may also be integral to the emergence of flow experiences.

The concept of flow may also facilitate understanding why our task complexity manipulation did not produce its anticipated effects. In the present study, we manipulated task complexity with the intention of determining whether the effects of distortions in time perception (on mood and task enjoyment) would vary along with the complexity of the task being performed. This hypothesis was not supported. Mood and enjoyment were uniformly lower for participants in the high complexity task, regardless of the objective duration of the reading task. The same pattern held when participants were grouped according to whether they experienced time flying (although only very few participants in this group experienced time flying). The more complex task seemed to frustrate most participants.

Csikszentmihalyi has suggested that an optimal match between one’s skill level and the complexity of a task helps set the stage for flow experiences. The present study did not include any measure of participants’ pre-existing level of mirror-image reading skill level, so it is impossible to know if there was any mismatch between the task and participants’ skills. However, it is not unreasonable to speculate that participants might have been unwilling to deploy existing skills and might have objected to the difficulty of this experiment compared to others used to gain course credit. Given

FIGURE 5

Mean Enjoyment Scores Across Experimental Groups

![Chart showing mean enjoyment scores across experimental groups.](chart.png)

Note: Higher scores represent greater enjoyment of the reading task. N = 77.
this potential lack of motivation, it is not surprising that enjoyment scores were uniformly low in the high complexity group and that positive mood change did not emerge. This was even true in the 5-min condition, where the reading task was (objectively) much shorter than expected and most participants (9 out of 14) experienced time moving more quickly than usual. Although the small sample size was clearly an issue, providing participants with alternative forms of motivation (e.g., stimulating a testing situation with the promise of interesting information about oneself) might have lead them to engage with the reading task differently. In the end, however, the major reason for the lack of the hypothesized interaction was most likely the ineffectiveness of the task manipulation and a lack of power caused by our small sample size.

**Future Directions: Exploring the Dynamics of Flow**

Future research can shed further light on the importance of time perception as the linchpin in and clearest indication of flow. Are our time perception mechanisms integral for achieving the flow state? Is such a state possible in the presence of lesions to the brain areas responsible for the perception of time on the scale of minutes to hours? Can distortions in time perception indicate if someone is in a state of flow, or even in a state of boredom, or anxiety? This type of research might have lead them to engage with the reading task differently. In the end, however, the major reason for the lack of the hypothesized interaction was most likely the ineffectiveness of the task manipulation and a lack of power caused by our small sample size.

Our findings also suggested that time perception mechanisms may be involved in a wider system of processes that work together systemically to give shape to our stream of consciousness. Specifically there may be moment-to-moment fluctuations in immersion, time perception, enjoyment, and mood that were not captured by our study. In the present study, these factors were only measured at the end of the reading task. A design that incorporates constant monitoring of attention/distraction (e.g., following shifts in eye gaze) may help reveal more nuanced information about differences in how one commits cognitive resources toward an attention-demanding task across different levels of complexity. Such an approach may help to distinguish between two separate (but related) effects on time perception: (a) the effect of initial interest in a task on time perception early in a task, and (b) the effect of immersion and sustained focus on time perception as the task continues (and whether interest is required for the development of full-blown immersion).

This type of research would be especially interesting in light of how time perception changes along with task duration and complexity (see Figure 1). Participants in the 5-min high complexity condition experienced time passing fastest, with a significant slowing of time for their 10-min counterparts. Participants in this group also produced high immersion scores. This pattern (high immersion together with a speeding up of time) was unique to the high complexity group and is suggestive of a flow-like experience, which then dissipated under longer exposure to the high complexity task. This was accompanied by a number of trends that are theoretically suggestive. For example, for participants in the high complexity condition, distortions in time perception varied along with immersion. The same pattern did not hold for the low complexity group, where there were no differences in immersion across the task duration groups. The differences in these patterns suggested that task complexity may play an important role in the relationship between immersion and distortions in time perception, which we have demonstrated is predictive of positive changes in mood. Future research can address both the role of task complexity in the emergence of flow states and whether task complexity might influence the speed at which flow states emerge and dissipate.

In applied settings, it would be important to further understand how cognitive factors (e.g., task complexity, immersion, processing capability, attention) affect time perception. Slight manipulations in these cognitive factors may help to influence mood and performance in work and school contexts. In these contexts, task complexity would be an especially relevant variable that could be easily manipulated in an effort to tailor
our experiences and the experiences of those with whom we collaborate.

If, as Sackett and colleagues (2010) have suggested, individuals use perceptions of the flow of time as a metacognitive cue used in framing an event, it may be possible to look forward to a day when we can reshape our everyday experiences by manipulating the subjective flow of time. The present study advanced the overall understanding of time perception and its connection with subjective experience of an event. It seems that, when time passes quickly, people have a more positive experience. Further, we replicated work on how time perception is constructed. We also furthered the understanding of the role that time perception plays in a flow experience and that immersion in a task significantly contributes toward a positive experience only if time perception has been altered accordingly. Understanding how to manipulate time perception to a desirable rate may be an important step in improving everyday life. Our work may suggest that time perception is a malleable intrinsic motivator, which people continuously seek to shape in a satisfying manner. Although people are not always conscious of their perception of time, time perception does affect the way that people feel about their experiences.

References


Author Note. Matthew Pageau, Department of Psychology, Saint Joseph’s University, PA; Seth Surgan, Department of Psychology, Worcester State University, MA.

Correspondence concerning this article should be addressed to Matthew Pageau, Saint Joseph’s University, Philadelphia, PA 19131. E-mail: mpageau1@worcester.edu
Think About It: Cognitive Load and the Trolley Scenario as an Analogue of Gun Violence

Molly Ramsden
Castleton State College

ABSTRACT. Moral dilemmas present an opportunity to examine the impact of distance on the acceptability of violent actions. The high-distance trolley and low-distance footbridge scenarios contrast low- and high-distance killing situations. Using these dilemmas, I tested the mitigating effect of distance on the acceptability of violence. Following a cognitive load task, decision-making centers in the brain switch from rational to emotional. Using the high-distance trolley and low-distance footbridge scenarios, I tested the impact of cognitive load and distance on the acceptability of the decision to kill 1 person, saving 5. When cognitive load increased, the acceptability of the decision to kill decreased in the trolley scenario, but not in the footbridge scenario, $F(1, 106) = 4.68, p = .033, \eta^2_p = 0.042$. Results suggested that, by depleting the rational brain with the cognitive task, the emotional brain made the decision not to kill. This research may be extended to gun violence because the presence of a gun typically places a shooter at a higher distance from a victim, which is analogous to the distance in the trolley scenario. Consequently, as in the trolley scenario, the acceptability of gun violence may be reduced as cognitive load is increased.

Multiple studies have assessed moral reasoning through the trolley and footbridge scenarios. These scenarios involve having to choose between saving five lives and saving one life. This problem is posed to the participant in two different contexts. In the high-distance trolley scenario (Foot, 1967), the participant imagines pulling a switch, which will result in saving the lives of five men, while leading to the death of one man who is standing on the tracks at some distance away from the participant. In the low-distance footbridge scenario (Thomson, 1985), the participant imagines sacrificing the life of a man by pushing him off a bridge to save the lives of five other men standing on the tracks below. Typically, participants are more willing to throw the switch in the trolley scenario and sacrifice one life to save five than they are to push the man off the bridge in the footbridge scenario to achieve the same utilitarian outcome of sacrificing one to save five.

Navarrete, McDonald, Mott, and Asher (2012) elaborated on the trolley scenario by adding a second condition. In the first condition, the participant had to pull the switch to achieve the utilitarian outcome of killing one to save five. In the second condition, the switch had already been pulled, and so the utilitarian result of killing one to save five would be achieved without participants’ involvement. Participants who had to pull the switch experienced significantly higher signs of distress than those who could have achieved the same utilitarian outcome by doing nothing.

Navarrete et al. (2012) also increased the ecological validity of their study by designing a virtual-reality trolley scenario. The trolley scenario is often given in a written transcript format. In this...
format, approximately 90% of participants report that they would pull the switch. However, it may be easier for individuals to say that they would pull the switch on a questionnaire when they are isolated from the reality of the situation than to actually pull it. In the virtual reality enactment, participants had to physically pull the switch to reach the utilitarian outcome. Navarrete et al. (2012) found that approximately 89% of participants were willing to pull the switch.

Bleske-Rechek, Nelson, Baker, Remiker, and Brandt (2010) used two studies to analyze the impact of sex, age, romantic involvement, and genetic relatedness on the willingness to act in the trolley scenario. They found that the higher the relatedness of the victim, the less likely participants were to pull the switch to sacrifice them. They found the opposite for age: the younger the victim, the less likely the participant was to pull the switch. When the victim was a romantic partner, both men and women typically refused to pull the switch.

Gikara, Farnsworth, Harris, and Fiske (2010) supported these results in their investigation of the effects of the victim’s in-group or out-group status relative to the participant. Participants were more likely to sacrifice an out-group member than an in-group member. Shallow, Iliev, and Medin (2011) examined the effect of the number of deaths caused by the participant’s utilitarian action. They had four levels that each participant evaluated: push intervention, where the five could be saved by pushing a single person off a bridge; switch intervention, where the five could be saved by redirecting the trolley onto a different track, but two people die; switch intervention, where the five could be saved by redirecting the trolley onto a different track, but four people die; and omission, where doing nothing indirectly caused five people to die. Pushing was viewed as the most difficult decision even when pulling the switch caused four deaths. There was no significant difference between killing two with the switch and killing four.

This created what Cushman and Greene (2012) called the dual process model. Moral dilemmas create tension between two distinct processes: the rational choice of saving the many versus the emotional choice of not wanting to kill. No matter the choice, part of the brain will be dissatisfied. The above research has helped to demonstrate this. In Bleske-Rechek et al.’s (2010) study, the added variable of relatedness tipped the scale in the dual process model. It increased the emotionality of the decision, thus exhausting rationality. This allowed for the deontological result that, the more related the victims were to the participant, the less likely the participant was to be willing to pull the switch and sacrifice them. This is true for many moral dilemmas. Rational and emotional processes are at a stalemate until other factors (relatedness, number of victims, methodology, and especially distance from victims) tip the scales.

The dual process model can be thought of as similar to a camera (Greene, 2014). There is an automatic mode (emotional/affective) and a manual mode (rational/cognitive). The automatic mode requires little thought and exists as a fallback when the manual mode, which can be difficult in conflicting situations such as moral dilemmas, becomes exhausted. A neurological basis for this theory shows that activation in the ventral striatum and the ventral medial prefrontal cortex was associated with the automatic/affective processes, and the dorsolateral prefrontal cortex was associated with the manual/cognitive processes (McClure, Laibson, Loewenstein, & Cohen, 2004).

Scans using fMRI have shown increased brain activity in regions dealing with emotions for close proximity or more personal moral dilemmas such as the footbridge scenario. These brain regions include the medial prefrontal cortex and the amygdala (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Participants in the footbridge scenario who were more willing to perform the utilitarian action showed increases in brain regions associated with cognitive control including the dorsolateral prefrontal cortex and regions in the inferior parietal lobe (Bunge & Wallis, 2007; Miller & Cohen, 2001). This was supported by Bartels (2008) who found that individuals with a more rational and less intuitive thinking style were more likely to make utilitarian judgments and by Hardman (2008) who found that individuals who scored high on a cognitive reflection test were approximately twice as likely to approve the use of a human trolley stopper in the low-distance footbridge scenario. The cognitive reflection test asked participants questions such as, “A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?” It aimed to assess the impulsive versus reflected-upon response (Frederick, 2005). The impulsive response would be 10 cents, but with reflection, the correct response was realized to be 5 cents.

Because it was clear that cognition played a role in moral dilemmas, it was important to understand
the factors associated with the decision-making process. Shiv and Fedorikhin (1999) investigated affective and cognitive processes in decision making. They argued that the cognitive (rational) processes could be distracted, at which point decision making would revert back to affective (emotional) processes. This phenomenon was illustrated in their 1999 study where participants who repeated a seven-digit string of numbers (high cognitive load with low cognitive resources) chose chocolate cake over fruit salad; but participants who repeated a two-digit string of numbers (low cognitive load with high cognitive resources) chose fruit salad over chocolate cake. After they recited the memorized string of numbers to the researcher, participants made their decision as part of a “reward” for participating in the study. Shiv and Fedorikhin argued that participants under high cognitive load had fewer cognitive resources left to allocate to decision making, allowing the affective processes of the brain to make the unhealthy choice, whereas participants under low cognitive load made the healthy choice because they could allocate cognitive resources to the task at hand.

Shiv and Fedorikhin (1999) suggested an interactive organization of decision-making processes in the brain, in which cognitive decision-making processes usually predominate. However, when these processes are exhausted and/or distracted, people fall back on affective processes. When cognitive load is low, and therefore resources are high, the cognitive processes maintain control.

This reasoning may help explain the results of Greene, Morelli, Lowenberg, Nystrom, and Cohen (2008), in which cognitive load increased judgment time in high-stakes moral dilemmas, but had little effect on the type of decision made (utilitarian vs. deontological). Green et al. (2008) looked at high-stakes moral dilemmas (footbridge scenario), which were shown to be conflict-inducing in participants in previous research. Greene and colleagues argued that cognitive processes govern utilitarian judgments and affective processes govern nonutilitarian decisions. Thus, increasing cognitive load by requiring participants to acknowledge fivees in a random sequence of numbers streaming across the top of the computer screen, in which the footbridge scenario is presented, increases judgment times. This is because the cognitive load task depleted cognitive resources that were necessary in determining the morally acceptable decision.

The dual process model talks of conflict between the affective and cognitive processes when deciding on moral dilemmas (Greene et al., 2008). By applying Shiv and Fedorikhin’s (1999) interactive model of cognitive processes, a possible explanation can be formed. The footbridge scenario displays higher levels of affective processes because the conflict involved in the dual process model interferes with the cognitive processes. Affective and cognitive processes are at a standstill, but the lack of distance adds an emotional component, thus straining and overpowering the cognitive resources necessary to make a rational utilitarian decision. This allowed the decision to be affectively influenced. Thus, when no other factor is involved, the footbridge scenario will show affective deontological results. This is further supported by Koenigs et al. (2007) who found that patients with brain lesions in the ventromedial prefrontal cortex, which is connected to emotional decision making (McClure et al., 2004), made surprising utilitarian judgments in moral dilemmas such as the footbridge scenario that tend to display deontological results. This implied that neural structures play a key role in deontological decision making.

By adding the cognitive load factor as Greene et al. (2008) did, cognitive processes were distracted and/or depleted, but because the norm for the footbridge scenario starts with cognitive processes distracted and/or depleted (from the internal conflict posed by the dual process model), it would be difficult to see a significant difference in choices. However, a difference was apparent in the length of time it took participants to make the choice. Going back to Greene’s (2014) metaphor, the brain became too exhausted from fighting with the manual mode of the camera, and so it reverted back to the automatic mode. The few who still chose the utilitarian act were probably those described by Bartels (2008), who naturally had a stronger rational way of thinking.

These studies also brought up the interesting point of what is happening to the cognitive processes that allows for the affective to surface. Shiv and Fedorikhin (1999) and Greene et al. (2008) were unclear about whether cognitive load tasks deplete cognitive resources or distract them. Greene et al. (2008) implemented the cognitive task simultaneously with the measurement of the dependent variables, in essence distracting the participants. The similarity in the results of Shiv and Fedorikhin (1999) when the cognitive task was implemented before the measure of the dependent variable provided evidence that depletion, as well as distraction, of cognitive resources...
occurred in both situations. If results were achieved solely with distraction, then cognitive resources would be restored (and a healthy choice would be made) once the cognitive task was completed (the number string recited). However, the apparent lag in the return of cognitive function (choosing the chocolate cake) implied that the resources were not immediately restored and were depleted instead. The combination of this depletion factor and the distance factor associated with the trolley and footbridge scenarios will extend our understanding.

Although much research has focused on the low-distance scenarios (footbridge), I focused on high-distance scenarios (trolley), to address the relative ease of distance killing. Grossman (1996) described this high-distance effect in his book, On Killing: The Psychological Cost of Learning to Kill in War and Society. He created a spectrum of weapons used in combat including their respective kill distances and provided evidence that the farther the killer is from their victim, the easier it is to kill. This could well be connected to events such as school shootings. Notably they are shootings and not stranglings or stabbings. This also has implications for military drones or bombers. How easy is it for someone sitting in a control center thousands of miles away from a war site to press one button that destroys an entire city within seconds? Grossman (1996) argued that, as distance increases, the resistance to kill decreases.

Weapons such as guns, drones, or bombs, similarly augment the distance from the victim as in the trolley scenario. The ease with which people pulled the switch or trigger can be unnerving. The goal of my research was to allow for a more thorough understanding of the cognitive and neurological processes involved in this high-distance decision making. By introducing a cognitive load task to the trolley scenario, I expected cognitive processes to be depleted. This would force participants to rely on affective processes, which were connected with lower acceptability of the utilitarian action of pulling the switch. Thus, this would lower the acceptability of killing from a distance.

In the present study, I hypothesized that there would be a main effect for distance. Participants in the trolley scenario condition would report higher acceptability of sacrificing one life to save five than those in the footbridge scenario condition.

In addition, I believed that there would be an interaction effect between distance and cognitive load. Participants in the trolley scenario condition with a higher cognitive load were expected to rate acceptability significantly lower than participants in the trolley scenario condition with a lower cognitive load. However, a significant difference between the high and low cognitive load conditions was not expected in the footbridge scenario condition.

Method

Participants
Participants were a convenience sample of 71 women, 33 men, and 6 unspecified (they chose not to disclose their sex; N = 110). Most were psychology students at a small rural liberal arts college in the Northeast. Participants received course credit for participation. Participants ranged from 18 to 42 years of age, with a median of 18. The mean GPA for participants was 3.2 (SD = 0.36). Most participants were Christian (51.4%) followed by nondenominational (44.1%), Jewish (2.8%), and Buddhist (1.8%). Data were not collected regarding ethnicity.

Measures
Modified versions of the trolley scenario (Foot, 1967) and footbridge scenario (Thomson, 1985) were used to assess the acceptability of aggressive acts. The high-distance trolley scenario involves pulling a switch to cause the utilitarian end and the low-distance footbridge scenario involves pushing a man to cause the utilitarian end. The participants read a scenario in third person featuring the actor named Sam. In each scenario, Sam killed the one to save the five, be that by pulling the switch or pushing the man. Third person was used, as opposed to first person, in an attempt to model previous studies (Schwitzgebel & Cushman, 2012). Participants rated the acceptability of the action taken in their scenario on a 6-point Likert-type scale ranging from 1 (extremely unacceptable) to 6 (extremely acceptable) to avoid neutral answers.

The computer program Dual N-Back (2011) was used to manipulate cognitive load on visual working memory. The Dual N-Back game field is a box separated into smaller squares by lines creating a 3 x 3 grid. During the game, an image is shown in one of the nine game squares. The participant must find a match between the current position of the image to the position shown one, two, or any number of steps earlier. If there is a match, the participant must press the corresponding match key (the letter A) on the keyboard. The more past steps the participant must remember, the higher the cognitive load. For purposes of this study, only Level 1, one step back (low cognitive load), and
Think About It | Ramsden

Level 2, two steps back (higher cognitive load), were used. The Dual N-Back is scored by percentage of correct responses (no response and a false positive are treated the same). The possible results range from 0% (none correct) to 50% (all correct). The scores do not range to 100% because the game has a built-in auditory working memory task that was ignored (the volume was turned off), accounting for the other 50%. The Dual N-Back has been shown to reduce the resources available for conscious and/or cognitive decision making, forcing decision making to rely on information provided by the unconscious and/or affective areas of the brain (Creswell, Bursley, & Satpute, 2013).

Procedure

After institutional review board approval (CSCIRB14: 05) was granted, participants were randomly assigned their cognitive load condition and distance (trolley or footbridge scenario) condition. They were taken into a computer lab where they completed the corresponding level of the Dual N-Back two separate times; totaling 42 to 48 trials (The game at Level 1 has 21 trials, Level 2 has 24 trials, thus 42–48 total). This would allow participants to become acquainted with the game. However, the two scores were averaged together so that any practice effect would be minimized.

To reduce the interval between the cognitive task (Dual N-Back) and the acceptability test (trolley or footbridge scenario), participants were asked to record their own scores on the Dual N-Back. The researcher monitored as the scores were recorded. Immediately after the Dual N-Back, participants completed the third-person trolley or footbridge scenario.

Results

A 2 (distance) x 2 (cognitive load) Analysis of Variance (ANOVA) was conducted to test the main effects and the interaction effect of distance and cognitive load on acceptability. The main effect for distance was significant, $F(1, 106) = 32.53$, $p < .001$, $\eta^2_g = 0.24$, supporting the first hypothesis in that participants in the high-distance condition ($M = 4.00, SD = 1.17$) scored acceptability higher than participants in the low-distance condition ($M = 2.61, SD = 1.43$).

The interaction effect between distance and cognitive load was also significant $F(1, 106) = 4.68$, $p = .033$, $\eta^2_p = 0.042$ (see Figure 1). This supported the second hypothesis that participants in the high-distance condition with high cognitive load ($M = 3.63, SD = 1.09$) would score acceptability lower than participants in the low cognitive load, high-distance condition ($M = 4.48, SD = 1.12$). Additionally, participants in the in the low-distance condition with high cognitive load ($M = 2.75, SD = 1.54$) scored acceptability higher than participants in the low cognitive load, low-distance condition ($M = 2.52, SD = 1.37$).

A binary logistic regression was used to further examine the predictive power of distance, cognitive load, and the interaction of distance and cognitive load on the binary split of the acceptability variable (1–3 = low acceptability, 4–6 = high acceptability). There was a significant difference in the probability to find the moral dilemma acceptable as distance, cognitive load, and the interaction between distance and cognitive load changed, $G(3) = 22.43$, $p < .001$. Distance was found to be the most significant in predicting acceptability ($\beta = 2.54$, $p < .001$). The interaction between distance and cognitive load ($\beta = -1.38$, $p = .11$), and cognitive load ($\beta = 0.61$, $p = .33$), were not significant, but showed similar trends to the ANOVA.

Discussion

The present study had two main foci: (a) to replicate the results previously shown by Greene et al. (2008) and (b) to extend these results by analogy to gun violence. As previously mentioned, Greene et al. (2008) focused mainly on scenarios that I have chosen to call low-distance scenarios. These are scenarios such as the footbridge scenario that tend to evoke strong emotions in participants, increasing decision-making difficulty. When cognitive load was introduced in Greene et al.’s (2008) study (simultaneous digit search task) and the present study (Dual N-Back task), little difference was seen in change of the acceptability of killing because few were willing to push the man off the bridge.

This can be explained through Shiv and Fedorikhin’s (1999) analysis of decision making. When resources are high (no cognitive depletion), people tend to make rational utilitarian decisions of choosing the many over the few, but when resources are low, people tend to make the automatic emotional deontological decision by allowing people to die without taking action. In these low-distance scenarios, decision-making difficulties stem from conflict between the automatic emotional and reflective cognitive, decision-making brain centers. This conflict depletes the cognitive brain’s resources as described by Cushman and Greene’s (2012) dual process model. When a
cognitive task is introduced (e.g., Dual N-Back), this merely depletes cognitive resources further. Thus, there was little change in the acceptability of killing the man on the bridge because decision-making processes were already being controlled by the automatic emotional processes of the brain. The key difference that cognitive load makes is in the high-distance trolley scenario. In the trolley scenario, cognitive resources are still relatively high because the greater distance between the killer and the victim reduces the emotionality involved. Consequently, there was a greater range of cognitive process depletion available in the trolley scenario, and this was the source of the interaction effect in my results.

Although many authors have focused on low-distance scenarios, I felt that more attention should be paid to high-distance scenarios due to their realistic similarities to real life situations in which people demonstrate an apparently callous disregard for human life. By agreeing with the utilitarian result created in these scenarios, participants are performing an action that involves taking another’s life. The ease with which participants pull a switch to take a life in the trolley scenario is disturbing to say the least. It is as disturbing as the apparent ease with which some people pull a trigger to take a life.

Extending this research by analogy to gun violence is important because it is not ethical to experiment on gun violence. Consequently it is vital to identify laboratory research paradigms that simulate factors associated with such violence. Analogues are difficult to identify, but the trolley scenario seems to have promise for representing an important dimension of the gun violence problem: distance between victim and killer. I wanted to investigate the power of the trolley scenario to elucidate this aspect of the gun violence problem.

Uniform Crime Report data from the Federal Bureau of Investigation provided another view of the relationship between distance and violence. These data showed how killing methodologies, which directly affect distance between victim and killer, are related to crime rates. The number of homicides committed each year is drastically higher with a gun than any other weapon. In 2011 alone, there were 8,583 homicides committed with a firearm, compared to 1,694 with a knife, or 728 with personal weapons such as hands or fists (“Murder Victims by Weapon, 2007–2011,” 2012). This, along with data collected from moral dilemmas such as the trolley scenario, has supported the theory that distance facilitates violent actions. The trolley scenario is not a perfect methodological match to situations involving gun violence, but it does provide a potentially useful analogue to important dimensions of this troubling social problem.

One other methodological concern has come to my attention. As mentioned earlier, participants completed a manipulation check for the effectiveness of the cognitive load variable by self-reporting their Dual N-Back scores. Self-reports can be influenced by social-desirability bias. I did not believe this would substantially influence my results for two reasons. First, participants were unaware of what their scores meant. The game frame that they were on did not explain the meaning of the reported numbers; it simply displayed them on the screen after the task was completed. Second, participants in the high-cognitive load condition reported very low Dual N-Back scores; they did not appear to be trying to make themselves look good. Consequently, I did not feel that self-reporting of Dual N-Back scores gave a false impression of the effectiveness of my cognitive load manipulation in this study.

There are several considerations I am making for future studies. To operationalize cognitive load, I chose a visual working memory task. This was due to past research that supported this model (Creswell et al., 2013; Greene et al., 2008). I now think it would be interesting to see how other working memory tasks may affect moral dilemmas. The Dual N-Back has the option of an auditory task as well as the visual one. Perhaps, by combining the two, a different level of cognitive load could be achieved. Also, by randomly assigning participants to different types of working memory tasks (visual,
auditory, combination, etc.) personal bias to one form or the other may be minimized. Some participants may be more adept at visual tasks, and thus would not feel the same cognitive load depletion effect as others who are less skilled in that area.

Although I focused mainly on decreasing utilitarian results in the present study, I think that there would be great utility in analyzing methods to overcome the affective influence in high-stakes moral dilemmas in order to increase utilitarian results. Glucose has been shown to reactivate the dorsolateral prefrontal cortex, increasing performance in bikers (Chambers, Bridge, & Jones, 2009). The dorsolateral prefrontal cortex is also one of the regions in the brain associated with utilitarian judgment (Bunge & Wallis, 2007; Miller & Cohen, 2001). By introducing a glucose variable into the model, it may be possible to increase utilitarian decisions in high-stakes dilemmas, further increasing our understanding of decision making within moral dilemmas.

My intention was to develop an analogue for investigating some of the factors associated with gun violence by demonstrating how the trolley scenario models the distance dimension of gun violence. In addition, I have demonstrated how understanding the factors influencing the utilitarian decision to kill are similar for the participant who throws the switch to save five individuals by killing one, and the individual with a gun who is distant enough from their target to make an intentional and rational decision to kill. I believe that the trolley scenario contains additional possibilities for explicating relationships associated with gun violence such as the need of police officers and soldiers to kill in the line of duty, and I expect to investigate these with future research.

References


Unpublished manuscript, London Metropolitan University, London, Great Britain.

Unpublished manuscript, London Metropolitan University, London, Great Britain.

Author Note. Molly M. Ramsden, Department of Psychology, Castleton State College, VT 05755. This research was funded by the Castleton State College Faculty-Student Research Grant with great support specifically from Terry Bergen. Correspondence concerning this article should be addressed to Molly M. Ramsden, Department of Psychology, Castleton State College, Castleton, VT 05755. E-mail: mmm01290@castleton.edu
The idea of the mad artist conjures images of a manic Vincent Van Gogh painting absinthe-fueled masterpieces between fits of sadness and rage (Falcone, Carlton, Janigro, Simon, & Franco, 2008) or a despairing Beethoven coping with a loss of hearing as he composes his greatest works as revealed in a letter titled the Heiligenstadt Testament by historians (Straus, 2006). The little known black dog bipolar disorder suffered by Winston Churchill made the organization and execution of Operation Bodyguard even more impressive (Storr, 1988). These individuals all suffered hardships as a result of a mental illness. The Heiligenstadt Testament, written by Beethoven to his brothers and discovered in his desk upon his death, revealed Beethoven’s struggle with doubt and anxiety that reached its breaking point with the loss of his hearing.

Some researchers have found positive correlations between trait anxiety and creativity (Carlsson, Wendt, & Risberg, 2000; Martindale, Anderson, Moore, & West, 1996). Others have concluded that

**ABSTRACT.** The purpose of the present study was to determine what effect facilitating and debilitating anxiety have on creativity when moderated by emotional states. Prior research into the relationship between anxiety and creativity has suggested that a 2-factor model should be applied to further study (Byron & Khazanchi, 2011). It was hypothesized that debilitating anxiety would be negatively associated with creative performance on a caption-writing task and that this relationship would be moderated by an emotional state. Participants in the study were undergraduates taking Introduction to Psychology (N = 102). After taking the 7-item Generalized Anxiety Disorder Scale (GAD-7) and a modified scale from Alpert and Haber (1960), participants were asked to write captions for 2 provided photographs drawn from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008). The 1st photo for all participants, a priming photo of either pleasant or unpleasant valance and high or moderate arousal, preceded a neutral photo moderate in valance and arousal. Statistical analysis of the 2 judges’ ratings on the creativity of the neutral captions yielded high interrater reliability. Although the findings did not support the hypothesis, facilitating anxiety significantly positively correlated to creativity (r = .27) and inversely correlated to debilitating anxiety (r = .40). Debilitating anxiety significantly correlated with anxiety as measured by the GAD-7 (r = .53). Findings suggested that facilitating anxiety, not debilitating anxiety, has a significant effect on creative ability as measured by the caption-writing task.
anxiety interferes with or is unrelated to creative performance (Byron & Khazanchi, 2011; De Dreu, Nijstad, & Baas, 2011; King, Walker, & Broyles, 1996). Are the men and women renowned as mad artists merely exceptional intellectuals, or does their creative ability lie dormant in all who suffer at the hands of anxiety, unrevealed to them until, under the right conditions, they find themselves capable of more complex inner thoughts and emotions? To get a little closer to the answer of this question, the present study sought to determine the relationship between creativity and anxiety when a two-factor model for anxiety was applied.

**Debilitating and Facilitating Anxiety**

Anxiety can have both debilitating and facilitating effects (Raffety, Smith, & Ptacek, 1997). Facilitating anxiety serves as a motivating factor, and debilitating anxiety has been associated with worry and distraction (Raffety et al., 1997). In a study examining the effects of an academic stressor test on academic performance, participants were asked to maintain diaries recording facilitating and debilitating anxiety in the days before an academic stressor. The researchers found that facilitating anxiety correlated positively with test scores significantly, and debilitating anxiety correlated negatively with test scores (Raffety et al., 1997).

**Creativity, Anxiety, and Arousal**

In a study examining the neurobiological factors of creativity, Carlsson et al. (2000) conducted a Creative Functioning Test (CFT) to test for creativity and measured anxiety using Spielberger state and trait anxiety inventories (STAI). The study revealed a significant correlation between trait anxiety in participants (found using the STAI) and a high score on the CFT, which indicated high creativity. Carlsson et al. (2000) also reported that overall blood flow during a resting period was significantly higher for highly creative participants. These findings reinforced those of Martindale et al. (1996), who found that participants high in creativity were more responsive to stimuli than were participants low in creativity. Martindale et al. (1996) also reported slower habituation of arousal among creative individuals than among those low in creativity.

Anxiety as it relates to creativity was examined in depth by a meta-analysis conducted by Byron and Khazanchi (2011). Byron and Khazanchi compiled 59 studies examining state or trait anxiety as it related to creativity. Results indicated that trait anxiety and creativity were negatively correlated. The significant positive correlations found between creative ability and trait anxiety in two studies (Carlsson et al., 2000; Martindale et al., 1996) and the more frequent reports of negative correlations (Byron & Khazanchi, 2011; King et al., 1996) have suggested that the relationship between these two factors may be moderated by other factors. King et al. (1996) examined the five factor model of personality in relation to creativity and found a lack of relationships among creative ability, creative accomplishment, and neuroticism (a personality factor highly correlated with trait anxiety).

**Creativity and Mental Health**

The relationship between depression and emotional expression and creativity has been found to be moderated by the emotional state of the participant (Forgerd, 2011). Lower levels of depression, measured to detect predisposed mood, were correlated with higher scores on a creative caption-writing and scoring task when induced by a negative emotion.

Depression carries a significant diagnostic overlap with generalized anxiety disorder (GAD; Moffitt et al., 2007). Symptoms of GAD, according to the Diagnostic and Statistical Manual of Mental Disorders, include excessive worry, difficulty in controlling such worry, and disruptions to daily routines as a result of anxiety and worry (5th ed; DSM-V; American Psychiatric Association, 2013). These symptoms are similar to those of debilitating anxiety as discussed by Raffety et al. (1997). For the present study, it was expected that the relationship between debilitating anxiety and creativity would be similar to that between depression and creativity.

In a study spanning 40 years, Kyaga et al. (2013) revealed that authors were more likely to suffer from anxiety disorders and other psychiatric disorders than were those from other professions. Authors were also more likely to suffer from unipolar depression and drug and alcohol abuse (Kyaga et al., 2013). In their meta-analysis, Byron and Khazanchi (2011) noted that future research “should consider the two-component model of anxiety in relation to creative performance as it may reveal possible differential relationships between different components of anxiety and creativity” (p. 279). Byron and Khazanchi believed that the next step in studying anxiety as it related to creativity was examining creativity in the context of the two-factor model.
Hypotheses
For the present study, we first hypothesized that debilitating anxiety would be negatively associated with creative performance on a caption-writing task. Our second hypothesis predicted that facilitating anxiety would be positively associated with creative performance. We proposed a third hypothesis that the induction of a negative state would yield greater creative performance from participants, following a fourth hypothesis that the relationship between facilitating anxiety and creative performance would be moderated by the emotional state of the participant, being stronger when a negative emotional state has been induced in the participant.

Methods
Participants
Participants (N = 102) were recruited from the participant pool of the department of psychology at a midsized eastern university. No demographic information on the sample is available. However, demographic information on the university’s subject pool is available. Of the 700 participants who made up the subject pool for the psychology department, age ranged from 17 to 26 years (M = 18.76, SD = 1.00). The subject pool was made up of 58% female students, 41% male students, and 1% of students who declined to answer questions regarding sex. The subject pool was made up of 66% White, 20% Black, 4% Asian, 8% other students, and 2% of students in the subject pool who declined to identify a race. In the subject pool, 72% were first-year students, 20% were sophomores, 7% were upper-class students, and 1% declined to identify their class. In the subject pool, 12% of participants identified themselves as currently employed, and 26% indicated that neither of their parents had completed a college-level education.

Materials
The three measures employed in the present study for anxiety included the 7-item Generalized Anxiety Disorder Scale (GAD-7) borrowed from Spitzer, Kroenke, Williams, and Löwe (2006), and modified scales adopted from Alpert and Haber (1960) that measure facilitating and debilitating anxiety separately.

Debilitating and facilitating anxiety. Two filter questions were included in the modified Alpert and Haber (1960) measure as an attention check. Eight questions corresponded to a facilitating anxiety score, and nine items corresponded to a debilitating anxiety score. All items in the modified anxiety measurement contained a statement followed by a response range of (a) always, (b) most of the time, (c) not often, or (d) never. Test-retest reliability over a 10-week interval was .83 for facilitating anxiety and .87 for debilitating anxiety as reported by Alpert and Haber (1960). When reliability for the facilitating and debilitating anxiety measures was estimated in the current study, facilitating anxiety yielded a Cronbach’s α = .70, and debilitating anxiety yielded a Cronbach’s α = .72.

GAD. The GAD-7 was provided to participants to determine if and how much each participant experienced GAD. Participants responded to a prompt that asked them to identify the frequency that they experience seven symptoms of GAD over a 2-week period. Responses for participants range from (a) not at all sure, (b) several days, (c) over half the days, to (d) nearly every day. Responses were tallied, ranging from 0 to 3, and added up to create a score out of 21. Using this method, a higher score means that the participant suffers more from GAD, whereas a lower score implies that they suffer less. Reliability of the GAD-7 was found in the analysis of the measure performed by Spitzer et al. (2006) who reported a Cronbach’s α = .92. The present study reported reliability of the GAD-7 through a Cronbach’s α = .87.

Creativity. Ten photographs were selected for their valance (positive vs. negative emotional value) and arousal from the International Affective Picture System (Lang et al., 2008). Valance and arousal measurements for the 10 photos are listed in Table 1. Of the 10 photos, five categories were created, placing two photos in each category. The five categories included unpleasant moderate arousal (UM) for photos scored unpleasant in valance and moderate in arousal, unpleasant high arousal (UH) for photos scored unpleasant in valance and high in arousal, pleasant moderate arousal (PM) for photos scored pleasant in valance and moderate in arousal, pleasant high arousal (PH) for photos scored pleasant in valance and high in arousal, and neutral for photos scored moderately for valance and arousal. The neutral photo captions written by participants were then used as a measure of creative performance through the scores assigned to them by two judges.

Participants also completed a Self-Assessment Mannequin (SAM; Lang, 1980) after completion of the priming and neutral photo captions to measure valance and arousal levels in participants.
Responses to the SAM valence and arousal manikins have been validated by showing them highly correlated with response to valence and arousal semantic differential items (Bradley & Lang, 1994).

**Design and Procedure**

The present study employed a between-subjects design to prevent the influence of a carry-over effect during the creative generation task. The GAD-7 and modified Alpert and Haber (1960) measure were counterbalanced to rule out the effects of survey fatigue. Institutional review board approval (Protocol #UMCIRB 13-002029) was secured. Participants provided consent the day of testing before being provided the testing materials. After completing the two anxiety measures, participants were presented with one of the emotion-inducing photographs (UM, PM, UH, PH) and asked to write a caption related to the photograph. Within each experimental group (UM, PM, UH, PH), half of the participants received one photograph, and the other half received the other photograph. This ensured that the results found were not due to unique characteristics of a particular photo, but rather due to the valance and arousal levels of the photos.

After writing the caption for the first photograph, participants were asked to complete the SAM for valance and arousal. Following the SAM, each participant was presented with one of the two neutral category photographs and asked to write a caption related to that photograph. Following the neutral caption-writing task, participants were asked to complete another SAM set. For the photo set, participants were instructed to generate a caption for both photos that provided insight into the photo (see Appendix A). Participants were also told that they could be as creative as they wanted when writing answers and that no response could be wrong for this task. Participants were instructed to not revise their work once they had completed a caption and moved on to the next photo.

The intent of having participants first compose a caption for the priming photo was to induce the valance and arousal variable in the participant. The priming photos’ effectiveness was then checked by the SAM, after which it was expected that the manipulation would affect the caption task on the neutral photo. Priming photo captions were not scored because they were completed before any manipulation could be introduced.

After participants completed the creative generation task, two judges scored the neutral photo from each participant (see Appendix B). These scores were used as a measure of creative ability. Responses were indicated on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). For each statement, they answered seven questions including (a) “The caption is relevant to the photo”; (b) “The caption includes creative use of elements found in the photo”; (c) “The caption creates a story for what was happening in the photo”; (d) “In addition to creating a story for the photo, the story created was different from what one would expect for this photo”; (e) “The caption contains elements of self-expression”; (f) “The caption is of a good quality”; and (g) “The caption appears to be inspired by emotions or experiences.” The reliability of the creativity measure was estimated with the intraclass correlation coefficient, the value of which was .92.

**Statistical Analysis**

Linear correlation/regression analyses were employed to determine the nature of the relationship between creative performance and scores on the three anxiety scales. The G*Power program was employed to conduct power analyses. With \( \alpha = .05 \), the sample size of 102 provided 88% power for detecting a medium-sized correlation (\( \rho = .30 \)).

An Analysis of Variance (ANOVA) was employed to determine if the four experimental groups differed on creative performance. With \( \alpha = .05 \), the sample size provided 53% power to detect a medium-sized effect (\( f = .25 \)).

**TABLE 1**

<table>
<thead>
<tr>
<th>International affective picture system ID#</th>
<th>Photo valance arousal category</th>
<th>Valance</th>
<th>Arousal</th>
</tr>
</thead>
<tbody>
<tr>
<td>2375.1</td>
<td>UM–1</td>
<td>2.20</td>
<td>4.88</td>
</tr>
<tr>
<td>9415</td>
<td>UM–2</td>
<td>2.82</td>
<td>4.91</td>
</tr>
<tr>
<td>8120</td>
<td>PM–1</td>
<td>7.09</td>
<td>4.85</td>
</tr>
<tr>
<td>2340</td>
<td>PM–2</td>
<td>8.03</td>
<td>4.90</td>
</tr>
<tr>
<td>6260</td>
<td>UH–1</td>
<td>2.44</td>
<td>6.93</td>
</tr>
<tr>
<td>2811</td>
<td>UH–2</td>
<td>2.17</td>
<td>6.90</td>
</tr>
<tr>
<td>5621</td>
<td>PH–1</td>
<td>7.57</td>
<td>6.99</td>
</tr>
<tr>
<td>8186</td>
<td>PH–2</td>
<td>7.01</td>
<td>6.84</td>
</tr>
<tr>
<td>7497</td>
<td>neutral–1</td>
<td>5.19</td>
<td>4.97</td>
</tr>
<tr>
<td>7640</td>
<td>neutral–2</td>
<td>5.00</td>
<td>6.03</td>
</tr>
</tbody>
</table>

*Note: UM = unpleasant moderate arousal; PM = pleasant moderate arousal; UH = unpleasant high arousal; PH = pleasant high arousal; neutral = moderate pleasantness moderate arousal.

Valance and arousal measurements were drawn from research from Lang et al. (2008) on the International Affective Picture System.
To test the moderation hypothesis, within-group slopes for the relationship between creative performance and anxiety were compared using the methods described by Weaver and Wuensch (2013). This test involved computing the standardized slope (r) for predicting creative performance from level of facilitating anxiety within groups (one primed with a pleasant photo and the other primed with an unpleasant photo) and then employing Fisher’s z to test the null that rho was the same in both populations.

Results
A statistical analysis focused on judge interrater reliability, the relationship between creative performance and scores on the anxiety scales, and assessed whether the priming photos had an effect and the nature of that effect for both valance and arousal. An assessment of the relationship discovered by Forgeard (2011) between low levels of depression and higher scores on a creative generation task under the influence of an induced negative emotion and the present study was made to determine if anxiety fit into the model suggested by Forgeard (2011).

Captions (N = 102) were submitted to the judges to be scored. Each caption received two scores: one from the first judge, the other from the second judge. Interrater reliability, estimated using a two-way mixed, consistency, average-measures intraclass correlation coefficient (McGraw & Wong, 1996), was excellent, ICC = .92 (Cicchetti, 1994). A measure of creative ability was created through the averaging of both judges’ scores on captions.

The first two hypotheses were supported: Facilitating anxiety was significantly positively correlated with creative performance (r = .28, p = .005) and significantly negatively correlated with debilitating anxiety (r = -.40, p < .001). Debilitating anxiety was significantly positively correlated with anxiety as measured by the GAD-7 (r = .53, p < .001).

As a manipulation check, the effectiveness of the priming photos inducing their intended effects on participants was assessed through responses on the SAM for valance and arousal. The SAM response range for both conditions was between -4 and 4. Mean valance levels on the first application of the SAM were significantly higher when participants had been primed by a pleasant photo (n = 51, M = 1.31, SD = 1.66) as opposed to when they had been primed with an unpleasant photo (n = 50, M = 0.10, SD = 1.84), t(99) = 3.48, p = .001. The two pictures used in the unpleasant condition did not differ significantly from each other (M₁ = 0.08, M₂ = 0.12, SD₁ = 1.85, SD₂ = 1.88), t(48) = 0.08, p = .94. Likewise, the two pictures used in the pleasant condition did not differ significantly from each other (M₁ = 1.04, M₂ = 1.54, SD₁ = 1.99, SD₂ = 1.32), t(49) = 1.06, p = .30. No significant difference was detected in SAM arousal ratings between moderate (n = 45, M = -0.98, SD = 1.96) and high arousal priming conditions (n = 47, M = -0.57, SD = 2.21), t(90) = 0.92, p = .36. The two pictures used in the moderate arousal condition did not differ significantly from each other (M₁ = -0.05, M₂ = -0.72, SD₁ = 2.06, SD₂ = 2.44), t(43) = 0.98, p = .33. Likewise, the two pictures used in the high arousal condition did not differ significantly from each other (M₁ = -0.43, M₂ = -0.38, SD₁ = 2.21, SD₂ = 2.14), t(45) = 0.09, p = .92.

The second application of the SAM followed participants’ generation of the neutral caption. To determine if the priming effect persisted beyond generation of the neutral caption, SAM scores were compared between groups. Participants scored significantly higher on the valance scale of the SAM when they had experienced the pleasant priming condition (n = 47, M = 1.23, SD = 1.40) than when they had undergone the unpleasant priming condition (n = 50, M = 0.52, SD = 1.71), t(95) = 2.25, p = .027. Participants’ SAM ratings of arousal in the second application did not differ significantly between those primed with a moderate arousal photo (n = 45, M = -0.42, SD = 2.28) and those primed with a high arousal photo (n = 47, M = -0.40, SD = 2.15), t(90) = 0.04, p = .97.

The effects of priming on creative performance were tested with a 2 (high vs. low valance) x 2 (high vs. moderate arousal) ANOVA. The main effect of valence and the interaction fell well short of statistical significance (p = .29). Accordingly, the third hypothesis was not supported. Interestingly, despite an absence of effect of the arousal condition on participants’ SAM arousal ratings, creative performance was significantly higher in the high arousal condition (n = 54, M = 26.42, SD = 8.81) than in the moderate condition (n = 48, M = 22.32, SD = 9.14), F(1, 98) = 5.36, p = .023, η²p = .052.

A comparison of the relationship between facilitating anxiety and creativity between participants who were primed with a pleasant photo (r = .27) versus an unpleasant photo (r = -.31) was conducted to determine if the unpleasant condition would yield similar results to what Forgeard (2011) found in the negative emotional induction condition. We anticipated that the...
strength of the association between facilitating anxiety and creative performance would be greater when participants were primed with an unpleasant photo. No significant difference in strength of association was found between pleasantness conditions ($z = 0.24, p = .81$). Accordingly, Hypothesis 4 was not supported. An exploratory analysis indicated that arousal priming did not significantly moderate the relationship between creative ability and facilitating anxiety ($z = -0.27, p = .78$). When debilitating anxiety or GAD-7 anxiety were substituted for facilitating anxiety, the differences in strength of association remained insignificant.

### Discussion

The hypothesis that debilitating anxiety would have a negative effect on creative ability, and that this effect would be moderated by the emotional state of the participant, was not supported. A significant relationship between facilitating anxiety and creativity was observed, but the strength of this association was not significantly affected by differences in photo priming valances, revealing that emotional state did not act as a moderator in the observed relationship.

### Creative Ability, Two-Factor Model of Anxiety, and Happiness

Although the hypothesis that debilitating anxiety would be negatively correlated with creative ability was not supported, one type of anxiety was found to have a significant relationship with creative ability. Facilitating anxiety was positively correlated with creative performance. Although unexpected, this finding was consistent with the two-factor model of anxiety.

The assessment of the SAM valance category and photo priming types revealed a significant relationship with photo valance categories. Participants who were primed using a pleasant photo were more likely to score higher on the SAM valance category than their unpleasantly primed counterparts. Although this finding could support the effectiveness of the International Affective Picture System (Lang et al., 2008), future research should examine if this relationship was the result of differences in the valance of the captions generated by participants.

Valence of the priming photo had no significant effect on creative performance. This finding revealed that the happiness state of the participant does not act as a moderator in the observed relationship between facilitating anxiety and creativity.

### Creative Ability and Arousal

Although the condition of arousal was included to control for the possibility of an uncontrolled variable obscuring results, a significant finding resulted from its inclusion in the study. High arousal priming was significantly associated with greater creative performance.

Both applications of the SAM failed to reveal a difference in arousal between high and moderate photo priming conditions. This led to the conclusion that (a) the effect of the arousal priming was not salient to participants, (b) the creative ability measure was more sensitive to arousal ratings than to the SAM, or (c) the manipulation of arousal was confounded with some unknown factor that enhanced creative performance. Research in this area should look to using physiological measures of arousal and creativity as measured by performance on a standardized test such as the Remote-Associations Test in order to maintain a tighter control of variables not possible with the caption task and SAM (Mednick, 1968).

### The Two-Factor Model of Anxiety

Although the expected inverse relationship between creative ability and debilitating anxiety was not found, facilitating anxiety and debilitating anxiety were negatively correlated. This finding fit well into the two-factor model of anxiety and strengthened the assumption that these two factors are not independent of each other.

The positive correlation between high scores on the debilitating anxiety scale and high scores on the GAD-7 supported the two-factor model for anxiety. This indicates that the type of anxiety measured by the GAD-7 can be characterized as debilitating, which is not surprising given that the GAD-7 was designed to help diagnose GAD, an anxiety disorder characterized by anxiety strong enough to disrupt day-to-day functioning in an individual, and the two-factor model of anxiety exists. Future studies into the two-factor model for anxiety should examine the nature of this relationship.

### Limitations

Upon arrival at the testing room, participants were promised that their participation would remain completely anonymous. In addition to ensuring that their identity remained anonymous, we promised that no identifying or personal information would be obtained. Unfortunately, a statistical analysis of how different demographics such as age, sex, race/ethnicity, and grade point average could
have played into the results cannot be reported.

We lacked access to physiological measures of anxiety and arousal such as the measures detailed from Carlsson et al. (2000). Access to physiological measures in place of the SAM could have strengthened the present study. The use of a self-assessment was the only choice we had to measure participant arousal immediately after both priming and neutral captions were written.

Conclusion
The present research has shown that facilitating anxiety, not debilitating anxiety, has a significant effect on creative ability, and that this effect is not moderated by an emotional state (pleasurable affect) in the participant. Although the anticipated outcomes were not observed, the study did succeed in establishing a relationship between the two-factor model of anxiety and creativity. This research developed a test of creativity and a procedure that could successfully prime a participant as indicated in the valence condition of both applications of the SAM. Results of the present study revealed a significant relationship between facilitating anxiety and creativity previously unobserved, and gave insight into the often underappreciated psychology of creativity. The relationship between facilitating anxiety and creativity should be tested using standardized tests for creativity in order to control for the variables that the current study failed to do.

The effect of facilitating anxiety on academic success found by Alpert and Haber (1960) and the effect on creative performance discovered in the present research revealed that its effect could be strong in other aspects of an individual’s life as well. Future research should investigate what other types of behavior may be affected by facilitating anxiety.

The present study succeeded in its purpose to shed more light on the complicated relationship between emotional states and creative ability. The study also revealed that high scores on debilitating anxiety correlate positively with scores on a measure for GAD. Although no single study could reveal the secrets of creativity entirely, this study did reveal a piece of the delicate relationship that creativity holds with anxiety. It is in the interest of future researchers considering a two-factor approach to anxiety to observe the findings reported in the present study.

It should be the goal of future research in the two-factor model to investigate the inverse correlation between facilitating and debilitating anxiety, whether it is possible through different forms of therapy to increase an individual’s facilitating anxiety, and whether their debilitating anxiety decreases as a result. This research could benefit individuals who suffer from high amounts of debilitating anxiety. Research could also investigate whether changing individuals’ placement on the two-factor model of anxiety would alter the amount of anxiety they experience as measured by the GAD-7. The clinical applications of such research could help therapists develop new methods of treatment for GAD.

References
followed prospectively to age 32 years. Archives of General Psychiatry, 64, 651–660. doi:10.1001/archpsyc.64.6.651

**Author Note.** Glenn B. Little and Karl Wuensch, Department of Psychology, East Carolina University, NC. Glenn Little has since graduated from the East Carolina University Psychology Department. Correspondence concerning this article should be addressed to Glenn B. Little, 5900 Thornwood Ct, Raleigh, NC 27613. E-mail: BranhamGL@gmail.com

### APPENDIX A

**Instructions for Participants**

For this stage of the experiment, you will be provided with two photographs. For each, please generate a caption for the photo. Your caption should provide insight into what you believe is happening in the photo, while providing a narrative-like backdrop for the photo. There are no wrong answers, so feel free to be as creative as you wish when interpreting the photo and its message.

There is no minimum length requirement. However, please limit your responses to only the space provided. When you have finished composing a caption for a photo, please set it aside and move on to the next one. Please do not revise your earlier work once you have already begun work on the second photo. Please try to utilize correct spelling and grammar in your captions (your work will not be judged for its grammatical or spelling errors). If this is not possible, please do not let it prevent you from writing, but write as clearly as possible.

### APPENDIX B

**Judging Instructions**

For each statement, indicate your degree of approval using a scale where 1 = strongly disagree, 2 = disagree, 3 = mildly disagree, 4 = neither agree nor disagree, 5 = mildly agree, 6 = agree, and 7 = strongly agree.

1. The caption is relevant to the photo. (If not, judges may delete this case.)
2. The caption includes creative use of elements found in the photo.
3. The caption creates a story for what was happening in the photo.
4. In addition to creating a story for the photo, the story created was different from what one would expect for this photo.
5. The caption contains elements of self-expression.
6. The caption is of good quality.
7. The caption appears to be inspired by emotions or experiences.
Romantic relationships are a significant part of people’s lives, which makes understanding them important. Understanding relationships begins by understanding the way they start: how people choose romantic partners. Evolutionary theory describes logical reasons why some personal characteristics are considered more romantically desirable, and influence dating choices and behaviors.

One important characteristic is physical attractiveness. Attractiveness impacts perceptions of personality traits, marriage prospects, career attainment, job selection, and overall happiness in life, so it certainly follows that it is a part of the mate selection process (Cash & Kilcullen, 1985; Dion, Berscheid, & Walster, 1972; Marlowe, Schneider, & Nelson, 1996). In fact, level of physical attractiveness has been shown to be the most influential factor in attraction to a potential partner (Prokosch, Coss, Scheib, & Blozis, 2009; Walster, Aronson, Abrahams, & Rottmann, 1966).

One way to explain the reasons why different characteristics such as physical attractiveness are romantically desirable is using principles of evolutionary theory. According to Buss (1989), the claims of evolutionary theory regarding human mate selection are based on parental investment and fertility. In the reproductive process, women have a longer investment than men. A man must only invest the time needed for intercourse. A woman must invest a minimum of nine months before birth, and the woman must invest years in raising the child if the child is to survive. A man’s investment does not necessarily end after conception; he may defend, provide resources, and assist with childrearing. This discrepancy in amount of investment is why evolutionary theory asserts that women are more attracted to men who have traits such as industriousness and ambition that are associated with resource acquisition, and women are more...

ABSTRACT. The present research sought to test the relationship between career-oriented information such as college major and perceived romantic desirability. The research tested 3 hypotheses, all based on evolutionary theory. The first was that people with lower earning potential would receive lower desirability ratings than people with higher earning potential. The second was that more career information presented with the picture would increase the difference seen in the first hypothesis. The third was that career-oriented information would have a larger impact on ratings given by women than by men. A pilot study was performed to select the pictures with the most average level of attractiveness to include on the stimulus sheet. Only the first hypothesis was supported, and it was supported for three dimensions of romantic desirability: short-term relationship desirability, \( F(1, 111) = 5.66, p = .019, \eta_p^2 = .049 \), long-term relationship desirability, \( F(1, 111) = 4.22, p = .042, \eta_p^2 = .037 \), and physical attractiveness, \( F(1, 111) = 9.70, p = .002, \eta_p^2 = .080 \).
selective overall when choosing sexual partners. Although women are more selective overall, men place more value on level of physical attractiveness (Schwarz & Hassebrauck, 2012). Because women are fertile for a shorter period of time than men, it is more important for men to value reproductive capacity when searching for a potential mate. The outward indicators of reproductive capacity that men use when searching for a mate are youth and physical attractiveness (Feingold, 1990).

These sex-specific differences in characteristics that are desirable in a mate do not seem to be based on societal factors. Buss (1989) compiled 37 samples from 35 countries on six continents and five islands; he found that women more heavily valued financial capacity in 36 out of 37 samples of different cultures, and women valued traits of ambition and industriousness in 29 of 37 samples. In all 37 cultures sampled, he found that men valued physical attractiveness or good looks more than women did when searching for a potential mate.

Evolutionary theory rationalizes the reasons why specific characteristics are considered attractive. For example, physical attractiveness is an indicator of health (Buss, 1989). Specifically, indicators of facial attractiveness, symmetry, and perceived masculinity or femininity act as a proxy for health (Burke, Nolan, Hayward, Russell, & Sulikowski, 2013; Grammer & Thornhill, 1994; Little, Jones, & DeBruine, 2011). According to Little et al. (2011) facial asymmetry has been linked to physical problems, so symmetry indicates health. Thus, those characteristics that are found to be physically attractive are indicators of physical health. Another example is the importance of intelligence in men. Prokosch et al. (2009) found that intelligence as measured by the Wechsler Adult Intelligence Scale-III vocabulary subtest was significant in predicting appeal in both short-term and long-term relationships. Intelligence has been linked to both good genes and earning potential. Thus, women found intelligent men more attractive because they would likely produce intelligent offspring and would also have the capacity to be a provider.

However, the influence of intelligence on desirability varied based on the context; intelligence was not as important for short-term relationships as for long-term relationships (Prokosch et al., 2009). Other determinants of desirability such as creative intelligence, arrogance, and faithfulness have also been shown to vary in importance depending on whether the context was a short-term or long-term relationship (Gangestad, Garver-Apgar, Simpson, & Cousins, 2007; Haselton & Miller, 2006). Buss and Schmitt (1993) found that mate selection strategies varied based on the context of a long-term or short-term relationship. The context varied the adaptive problems faced by men and women. For example, long-term relationships benefit men because they can ensure paternity, and long-term relationships benefit women because they continue to have access to the man’s resources. This variation in motivations suggests that context should be taken into account when determining romantically desirable characteristics.

General assertions of evolutionary theory have received much support, but some research has found that aspects of evolutionary theory do vary across different individuals or populations. In a meta-analysis, Feingold (1990) found that type of research paradigm affected the size of the trends predicted by evolutionary theory. He found that, in all five paradigms studied, men valued physical attractiveness in a partner more than women did, but that the difference was stronger in self-report methodologies than in those measuring actual social behavior. These findings still supported evolutionary theory, but brought into question the magnitude of the theory’s effects. There are nonlinear sociodemographic differences in effect size and evolutionary predictors of attractiveness such as a peak of increased attractiveness for a man at an income of $100,000 and racial differences in women’s willingness to marry a man who is unattractive or unemployed (Kenrick, Sundie, Nicastle, & Stone, 2001; Sprecher, Sullivan, & Hatfield, 1994). Despite their variation, all of these studies (Feingold, 1990; Kenrick et al., 2001; Sprecher et al., 1994) have shown support for the general trend of women valuing earning potential and men valuing physical attractiveness in a potential mate.

In contrast, White (1980) showed that, for people who had a romantic partner, similarity of attractiveness rather than level of physical attractiveness was related to relationship duration and progress. This finding supported a version of the matching hypothesis first introduced by Walster et al. (1966). The original intent of the hypothesis was that individuals prefer to engage in a romantic relationship with others who have similar social desirability, but research has found that people typically want to be in a romantic relationship with the most desirable potential partner (Taylor, Fiore, Mendelsohn, & Cheshire, 2011). This difference in desire for a partner with a high level of physical
attractiveness and actual behavior reflecting the importance of similarity is parallel to the difference in self-report measures and actual behaviors found by Feingold (1990). Both could be explained by a person's mate value. People have an idea of their own value as a mate, and this value determines what people look for in a romantic partner (Tadinac & Hromatko, 2007). If people believe that they have a high mate value, they are much less likely to consider engaging in a relationship with a less desirable potential partner than people who have a low mate value; this effect could be partially due to fear of rejection in pursuing a potential partner who is significantly more desirable (Taylor et al., 2011). Therefore, although everyone prefers the most socially desirable partner (White, 1980), self-assessed mate value acts as a moderator that supports similarity in attraction (Taylor et al., 2011).

Despite the differing results, the matching hypothesis can be considered within evolutionary theory. According to Buss and Shackelford (2008), people assess their own mate value and use this value to determine traits required of a potential partner. For example, women with a higher mate value have higher standards across multiple different indicators of desirability (good genes, good investment, good parenting, and good partner) than women who have a lower mate value.

The present study extended the assertions made by evolutionary theory by more closely defining criteria that indicate earning potential. Education has been used to represent earning potential (Stevens, Owens, & Schaefer, 1990), so the present study extended the representation to specific educational training, defined as college major. Because there is a vast discrepancy between the incomes of several types of professionals who all have the same level of education, this indicates that type of education is a more specific proxy for earning potential, which is the variable of interest. Although this specificity could potentially make college major a better proxy for earning potential than education level, no previous research has used major, so whether participants would make inferences about earning potential with major as the only source of information was unknown. Amount of information was included as an independent variable because varying the amount of information given to participants allowed the researcher to determine whether including additional leading information such as potential income would be necessary for participants to infer earning potential or whether major alone was sufficient. Additionally, physical attractiveness has been shown to affect romantic desirability, so a pilot study was conducted to select the most average faces for the main study so that level of physical attractiveness could be controlled.

The primary study tested three main hypotheses regarding potential effects of earning potential, sex, and amount of information on romantic desirability. Romantic desirability was divided into three dependent variables: physical attractiveness, short-term relationship desirability, and long-term relationship desirability. All three dependent variables were defined as a score from 1 to 5 taken from items on the survey. Earning potential was divided into two categories of high and low. Low was defined as the potential career having one of the four lowest mean annual incomes, which ranged between approximately $33,000 and $41,000, and high was defined as the potential career having one of the four highest mean annual incomes that did not exceed the upper limit of $100,000 found by Kenrick et al. (2001). The incomes of the high earning potential category ranged from approximately $86,000 to $97,000. Amount of information was used to determine whether participants tended to give the same ratings if they were only provided with information about major as when they were also provided with career goals and mean income for the potential career. This was a between-subjects manipulation defined by which of three versions of the survey, each varying amount of information given, the participant received. The first hypothesis was that information representing lower earning potential presented with a picture would cause desirability ratings to be lower than information representing higher earning potential. The second hypothesis was that more career information presented would increase the difference seen in the first hypothesis. It is unclear whether major was associated with future socioeconomic status, so the additional career information was expected to make the association more explicit and create a larger difference in ratings. The third was that career-oriented information would have a larger impact on the ratings given by women than by men, which would be consistent with evolutionary theory.

Pilot

Method
Participants. Students in two introductory psychology courses were recruited. All of the students in both classes elected to participate, and there were
a total of 42 participants. Students were not asked to disclose their sex to avoid any discomfort regarding sexual orientation because they were asked to indicate and then rate the pictures of the sex that they found most romantically attractive. Of the 42 participants, 24 rated the male pictures and 18 rated the female pictures. Because participants were recruited in an introductory course, most were younger undergraduates: 62% were first-year students, 12% were sophomores, 5% were juniors, 2% were seniors, and 19% were other or did not indicate a class level. The mean age was 18.56 years ($SD = 1.01$). Sixty-two percent were White, 14% were Black, 7% were Asian, 7% were multiracial, and 9% were a different race or did not provide a response.

**Materials.** An informed consent form was given to all participants. The informed consent contained a general description of the tasks required of participants, the voluntary nature of the research, the use of the results, and contact information for further questions.

Thirty-two male and 32 female pictures of faces obtained from the database of faces developed by the Artificial Intelligence Laboratory of the Faculty of Industrial Engineering in São Bernardo do Campo, São Paulo, Brazil (Thomaz & Giraldi, 2010) were used. The pictures used to judge attractiveness were faces rather than full body pictures because facial attractiveness is the first indicator used in judging attractiveness (Furnham, Laveney, & McClelland, 2001). Eight pictures of each sex were printed on each side of two sheets of paper. The pictures were numbered continuously across the pages (1–32 for each sex), and each picture had a height of approximately 2 in. and a width of approximately 1.4 in.

A survey containing 32 numbered 5-point Likert-type scales was used to measure attractiveness. The possible ratings ranged from 1 (very unattractive) to 5 (very attractive). Demographic questions at the bottom asked for age, race, class (first-year student, sophomore, junior, or senior), major(s), and minor(s).

**Procedure.** After institutional review board approval (2013-10-09 EXP) was given for both the main study and the pilot, the pilot study was conducted. The researcher was present at the beginning of the class time and distributed the informed consent forms. The informed consent form was explained, and the voluntary nature of the study was stressed. After questions were addressed, participants received the survey and the picture sheets.

They were instructed to choose to respond to either the male or female pictures, based on which they found most romantically attractive. Participants provided an attractiveness rating on the survey that corresponded to all 32 pictures (e.g., the first picture was numbered 1, and the participant would circle the rating for that picture next to the number 1 on the survey). After they completed the survey, they were given an opportunity to ask questions.

**Results**

The physical attractiveness scores from each of the pilot pictures were averaged, and the eight pictures for each sex that had an average closest to three, which was labeled average on the scale, were chosen. The average scores for the male pictures ranged from 1.83 to 3.08. The range of the pictures chosen was 2.30 to 3.08, and the mean rating of all of the chosen pictures was 2.57. The average scores for the female pictures ranged from 1.78 to 3.90. The range of the means of the pictures chosen was 2.67 to 3.35, and the mean rating of all of the chosen pictures was 2.97.

**Experiment**

**Method.**

**Participants.** In the same semester, participants were recruited from introductory psychology classes not used in the pilot, and from general required courses. These classes were selected because introductory psychology is one option to fulfill core requirements usually taken by first-year students and sophomores, and the general required courses are usually taken by juniors and seniors. The professors of these classes were asked for permission to use class time for students to participate. There were 122 total participants, but five were excluded for either not completing the survey or completing it in a way that could not be scored. Approximately 45% of the 117 were men and 55% were women. Participants had a mean age of 20.31 ($SD = 3.18$). Seventy-five percent of the sample identified as White, 14% Black, 6% multiracial, and 5% other races. Thirty-five percent of participants were first-year students, 16% sophomores, 29% juniors, and 20% seniors.

**Materials.** The eight male and eight female faces that were rated the most average by participants in the pilot study were made into a new stimuli sheet for the experiment. Because there was some variation in the scores of the faces rated most average, the eight faces for each sex were split into four sets of pairs of the lowest two scores, the next
lowest two scores, the next highest two scores, and the highest two scores. Each pair was split so that one went into the high earning potential group and one went into the low earning potential category.

Three forms of a survey all had the same four items to respond to for each face. Short-term relationship desirability was defined as the response to the second survey item: “Would you go on a date with this person?” Long-term relationship desirability was defined as the mean of the third and fourth items on the survey: “Could you see yourself in a long-term relationship with this person” and “Could you see yourself marrying this person?” Physical attractiveness was defined as the response to the first item on the survey: “Rate this person’s physical attractiveness.” The only difference between the three forms was the amount of information given about each of the faces. One form provided only the college major corresponding to each picture. Another form added career goals to major; the final form gave major, career goals, and mean salary of the intended career. Mean salary was obtained from the U.S. Department of Labor, Bureau of Labor Statistics (2013). Table 1 contains the career information included in the surveys. The surveys also included demographic questions asking for age, sex, race, class, major(s), and minor(s).

Procedure. All of the students were given an informed consent form. The consent form was explained, and there was time for questions before the beginning of the study. Each participant was given the same two sets of facial pictures, which were printed on the same page. They were instructed to choose to respond to either the pictures of either men or women, based on which they found most romantically attractive. The researcher distributed the survey forms so that the first participant received Survey 1, the second Survey 2, the third Survey 3, the fourth Survey 1, the fifth Survey 2, and so forth. Participants completed the surveys and returned them to the researcher. After all participants had finished, the researcher gave them another opportunity to ask questions and then thanked them for their time.

Results
Three 2 (sex: male, female) x 2 (earning potential: high, low) x 3 (amount of information: major; major and career goal; major, career goal, and salary) mixed Analyses of Variance (ANOVAs) were conducted using SPSS® with the two between-subjects (sex and amount of information) and one within-subjects (earning potential) independent variables. There was one ANOVA for each dependent variable: short-term relationship desirability, long-term relationship desirability, and physical attractiveness.

The ANOVA on long-term relationship desirability yielded only one statistically significant main effect of earning potential, $F(1, 111) = 4.22, p = .042, \eta_p^2 = .037$. The mean score for high earning potential was 2.19 ($SD = 0.84$), and the mean score for low earning potential was 2.06 ($SD = 0.69$; see Figure 1). The other two main effects and all interactions were nonsignificant; the nonsignificant $F$s ranged from 0.25 to 2.59, and the $ps$ from .07 to .78.

There were two statistically significant main effects using short-term relationship desirability as the dependent variable. One was the within-subjects manipulation of earning potential, $F(1, 111) = 5.66, p = .019, \eta_p^2 = .049$. The mean score for the high earning potential pictures was 2.89 ($SD = 1.16$), and the mean score for the low earning potential pictures was 2.66 ($SD = 0.83$; see Figure 1). The other significant main effect was sex, $F(1, 111) = 7.89, p = .006, \eta_p^2 = .066$. The men’s mean score was 3.01 ($SD = 0.88$), and the women’s mean was 2.57 ($SD = 1.06$; see Figure 2). The other main effect and all interactions were nonsignificant; the nonsignificant $F$s ranged from 0.17 to 0.83, and the $ps$ from .37 to .85.

Both earning potential and sex had statistically significant main effects using physical attractiveness as the dependent variable. The within-subjects manipulation of earning potential was significant, $F(1,111) = 9.70, p = .002, \eta_p^2 = .080$. The mean

<table>
<thead>
<tr>
<th>Major</th>
<th>This person plans to</th>
<th>Mean salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics</td>
<td>work in insurance as a financial analyst</td>
<td>$86,240</td>
</tr>
<tr>
<td>Biomedical engineering</td>
<td>work in medical equipment and supplies</td>
<td>$92,200</td>
</tr>
<tr>
<td>History</td>
<td>work for a local government as a historian</td>
<td>$37,430</td>
</tr>
<tr>
<td>Chemical engineering</td>
<td>work in synthetic fibers manufacturing</td>
<td>$96,110</td>
</tr>
<tr>
<td>Elementary education</td>
<td>work in a religious school as a kindergarten teacher</td>
<td>$40,920</td>
</tr>
<tr>
<td>Art</td>
<td>work as an independent artist</td>
<td>$33,670</td>
</tr>
<tr>
<td>Computer science</td>
<td>work in computer systems design</td>
<td>$93,340</td>
</tr>
<tr>
<td>Psychology</td>
<td>work in relief services as a social worker</td>
<td>$37,530</td>
</tr>
</tbody>
</table>

Note: The information included in each of the three surveys was cumulative; Survey 1 included the information from the first column, Survey 2 included information from the first two columns, and Survey 3 included information from all three columns.
score for the high earning potential pictures was 2.98 (SD = 0.65), and the mean score for the low earning potential pictures was 2.80 (SD = 0.72; see Figure 1). The other significant main effect was for sex, F(1, 111) = 13.18, p < .001, η² = .106. The mean score given by men was 3.06 (SD = 0.69), and the mean score given by women was 2.60 (SD = 0.67; see Figure 2). The other main effect and all interactions were nonsignificant; the nonsignificant Fs ranged from 0.17 to 2.40, and the ps from .10 to .84.

Two paired-samples t tests compared long-term and short-term relationship desirability. Focusing on only the four images associated with a high earning potential, the first revealed a statistically significant difference, t(16) = -7.30, p < .001. Participants gave higher ratings for short-term relationship desirability (M = 2.89, SD = 1.16) than long-term relationship desirability (M = 2.19, SD = 0.84). The same pattern was seen when focusing on the low earning potential images, t(116) = -13.05, p < .001. Short-term relationship desirability (M = 2.66, SD = 0.83) was again significantly rated higher than long-term relationship desirability (M = 2.06, SD = 0.69).

Discussion

The first hypothesis was supported across all three of the dependent variables; all participants tended to give higher overall desirability ratings to the pictures that were associated with higher earning potential. Although previous research documented a difference in mate preferences for short-term and long-term relationships (Prokosch et al., 2009), the current study found earning potential to be an important factor in both. The size of the effect may vary when other characteristics of the potential partner are considered in more complex models, but it seems that earning potential influences the three measured subdivisions of romantic desirability when very little is known about a potential partner.

Attractiveness has been shown to influence how people view others and choose a potential romantic partner (Cash & Kilcullen, 1985; Dion et al., 1972; Marlowe et al., 1996; Prokosch et al., 2009; Walster et al., 1966), but the present study showed that earning potential can affect ratings of physical attractiveness. Perhaps impressions are holistic, and positive attributes are associated with one another; people assume others who are more physically attractive have better traits, so if it is explicit that another person has a desirable trait such as high earning potential, the person might seem slightly more physically attractive. This is an idea that requires further investigation.

Neither the second nor the third hypothesis was supported for any of the dependent variables. There was no significant difference in desirability ratings based on how much information the participant received. This could have been because participants who were given less information made assumptions regarding possible income. Thus, all three groups might have been giving ratings based on roughly the same information, whether explicitly stated or inferred by the participant. This would imply that people either consciously or unconsciously make a judgment regarding romantic desirability based only on a person’s major. Thus, major alone is a sufficient proxy for earning potential in the absence of other relevant information. Alternatively, a flaw in this research could have failed to find a difference that actually existed.

There was also no significant interaction between sex and level of earning potential for any of the three dependent variables. Based on the assumptions of evolutionary theory, there should have been an interaction in which the preference for high earning potential was stronger for women than men because of a higher parental investment for women. Perhaps sex roles have changed enough since the original evolutionary research that there is an equal desire in both men and women for a partner with a high earning potential. Women are able to support themselves and a child without the help of a man. This conclusion could also be biased.

![FIGURE 1](image_url)

**FIGURE 1**

**Romantic Desirability Ratings Across the Within-Subjects Manipulation of Earning Potential**

Note: This figure illustrates the difference in the mean ratings for the high and low earning potential pictures in the three dependent variables.
by the fact that all participants were enrolled in a four-year college; perhaps educated women are more confident in their ability to support themselves. Additionally, men are taking on a higher level of parental investment such as staying home with children, which could lead men to look for a woman with higher earning potential.

Another explanation is that similarity, as described by the matching hypothesis, is more important than earning potential. Although the matching hypothesis can fit into an evolutionary framework, the criterion to determine desirability can include different characteristics than those expressed by evolutionary theory. Stevens et al. (1990) found that more attractive men were more likely to marry more attractive women, and that highly educated men were more likely to marry highly educated women, rather than an evolutionary view of different characteristics being more desirable for each sex (e.g., a man with a high earning potential would marry an attractive woman). Their conclusion stressed the importance of similarity in attraction, similar to the conclusion of White (1980). From the perspective of the matching hypothesis, participants in the present study could have used major to infer a similar level of intelligence or similar interests rather than a likely future income.

A finding that did not relate to any of the hypotheses was that men gave higher physical attractiveness and short-term relationship desirability scores than did women. In looking at the scores on the pilot attractiveness ratings, this finding for physical attractiveness is not surprising, but the reason behind it is unclear. When reviewing these findings in an evolutionary framework, it seems fitting that women would rate potential romantic partners lower than men do because they have the larger parental investment. This finding was similar to the results found by Schwarz and Hassebrauck (2012), supporting the idea that women have higher standards for a potential mate than do men. However, this trend was not supported using long-term relationship desirability as the dependent variable. Perhaps when long-term commitment is assumed, the difference in parental investment is no longer relevant. If the woman does not fear being left alone to raise a child because the relationship is long-term, there is not a justification for a sex difference. Alternatively, the faces that were chosen for the present study might not have been equivalent, and men do not generally rate possible partners as more attractive relative to women’s ratings. Another unexpected finding was that, across all the independent variables, the ratings for long-term relationship desirability were significantly lower than the ratings for short-term relationship desirability. However, in retrospect, it seems logical that people should be either more cautious or less certain when it comes to deciding on a long-term partner; long-term partnerships are a much larger commitment and usually require more information than was provided to participants. Both of these findings require further research to explain.

There were several limitations in the current study. The sample size was relatively small, which reduced the power of the statistical analyses, particularly for the interaction effects. A true difference resulting from manipulation of or interaction between the independent variables might have not had an effect large enough to be detected. In addition, the sample was taken from a small private Lutheran-affiliated institution, so the sample might not be representative of students in other universities.

Another limitation was that the design was relatively straightforward although actual relationships have many more complex factors. The study limited the description of the individuals to include only one major, but in reality some people have more than one major. The definition of high and low earning potential was based on salaries from careers related to each major, but all people are not able to or do not choose to work in a field related to their college major. For example, people who choose a major with very limited opportunities in related fields may either work in other areas (e.g.,

---

**FIGURE 2**

Statistically Significant Main Effects

![Graph showing statistically significant main effects](image)

*Note: This figure illustrates the difference in the mean ratings for men and women in two of the three dependent variables.*
food service or retail) or pursue further education in a different field. The analyses included major as a proxy for earning potential, but did not control for any other associations with majors. The difference might not actually have resulted from the different earning potentials represented by the majors, but some other associations such as personality characteristics that people may or may not tend to associate with certain majors. Also, the present study used self-report data, which tends to show a stronger effect than actual social behavior, so the results might not accurately reflect people’s typical behaviors (Feingold, 1990).

Another important consideration in understanding the findings is that only average or slightly unattractive pictures were used. Statistically speaking, most people are of average attractiveness, presumably lending more generalizability to our findings. However, the effects might have varied or the hypothesized interaction might have been present if extremely attractive or extremely unattractive pictures were used.

Another limitation that may be significant was that the study did not take sexual orientation into account. This was a deliberate decision because disclosure of sexual orientation can be a sensitive or difficult issue for some people. However, there may be differences in partner preferences between heterosexual people and gay men and lesbians. The author weighed the ethical concerns of excluding gay participants from the study and determined they outweighed the potential for a small percentage of gay participants to shift results if indeed they were answering in a systematically different way. Future research may take these variables into account to understand the unique impact of sexual orientation on the relationship between romantic partner desirability and earning potential.

The present study had several implications. The implication for college students was that students extrapolate earning potential from major. Because earning potential is related to relationship desirability, the simple act of telling a new acquaintance one’s major prompts the other person to make assumptions that affect one’s desirability. Whether this knowledge would actually change behavior (e.g., not disclosing major if associated with low earning potential or actually becoming a factor in choosing major) is a topic for further research.

A fundamental part of evolutionary theory was not supported due to the lack of an interaction between sex and earning potential. This should be addressed in future research to determine whether there has been a change that makes the current evolutionary view obsolete or whether the limitations of the present study created a type II error. If the interaction effect does not actually exist, future research should either further explore the idea of similarity as the best predictor of attraction (Stevens et al., 1990; White, 1980) or create a new model to explain romantic attraction.

The present study demonstrated that college major can act as a proxy for earning potential in much the same way as income acts as a proxy for earning potential. Earning potential is a significant characteristic that people use to determine romantic desirability. However, there were many limitations that should be addressed in future studies such as identifying covariates of earning potential to see whether it is actually the causal factor in the differences in desirability ratings. This research is needed in order to better describe the initiation of romantic relationships, which are an important part of many people’s lives.

References


**Author Note.** Amanda L. Young, Department of Psychology, Lenoir-Rhyne University, NC.

Amanda Young is now at Department of Psychology, North Carolina State University.

Special thanks to Amy M. Hedrick in the Psychology Department of Lenoir-Rhyne University for suggestions to the content of this document.

Correspondence concerning this article should be addressed to Amanda Young, Department of Psychology, North Carolina State University, Raleigh, NC 27695-7650. E-mail: alyoung6@ncsu.edu
Despite women’s economic, social, and political gains in recent decades, the association of women with housekeeping and caretaking responsibilities has stagnated. Throughout the world, women continue to report spending greater proportions of time on housework than men (Lachance-Grzela & Bouchard, 2010; Ruppanner, 2010). Because women are often stereotyped as communal and nurturing, they are often believed to have a natural propensity for household tasks such as cleaning and doing laundry (Bem, 1974). Such stereotypes have the potential to relegate women to domestic roles and constrain their behavior (Glick & Fiske, 2001).

Gender role stereotyping is prominent in media and advertisements wherein companies rely on socially prescribed expectations to sell their products (Infanger, Bosak, & Sczesny, 2011). When framed in the context of humor, such stereotypes may evade criticism and perpetuate prejudice (Ford, Boxer, Armstrong, & Edel, 2008). To date, several studies have examined the content
of gender-based stereotypes in the media as well as how these stereotypes influence perceptions of both women and men. For example, Eisend's (2010) meta-analysis illustrated the ways in which advertisements “mirror” gender role expectations within a particular cultural context, yet provided little evidence that stereotyped advertisements exacerbate prejudicial values. On the other hand, several researchers have indicated that exposure to stereotypes reinforces negative attitudes toward women and prevents advances in their social status (Heilman, 2001; Romero-Sánchez, Durán, Carretero-Dios, Megías, & Moya, 2009; Tiggemann & McGill, 2004).

Although some researchers have analyzed responses to sexist humor (Ford et al., 2008; Ford, Wentzel, & Lorion, 2001; Greenwood & Isbell, 2002), there has been a lack of research in the area of gender-based humor specifically within the context of advertising and product promotion. The present study attempted to fill this gap, focusing on responses to humor as a marketing strategy. Specifically, the following questions were taken into account: Upon hearing gender-based humor, do people tend to respond with agreement, dissent, or indifference? What arguments or rationalizations do people draw upon when confronted with gender role stereotypes? In light of the mirror perspective (i.e., that advertisements convey the gendered attitudes of a particular culture), these questions were believed to be particularly salient in that they reveal predominant perceptions regarding women’s roles in society (Eisend, 2010). Considering the evidence that stereotypes may exacerbate negative attitudes toward women, understanding the ways that people respond to this type of humor can lead to a greater understanding of the media’s role in the persistence of sexism.

Stereotypes and Humor in Advertisements

Advertisements may rely on stereotypes to compensate for a lack of time and space, providing viewers with the most basic information possible. Numerous studies have identified advertisements as a source of gender role stereotypes. In advertisements, women are often portrayed as attractive, young, nurturing, and childlike (Kim & Lowry, 2005). On the other hand, men are often overrepresented as authority figures and are more likely to be shown using logic (Monk-Turner, Kouts, Parris, & Webb, 2007). In addition, men appear in advertisements more often, and when women are portrayed, they are often sexualized or victimized (Stankiewicz & Rosseli, 2008). Stereotyped depictions of women in advertisements are often effective marketing strategies. For instance, Infanger and colleagues (2011) found that advertisements portraying women with communal traits (e.g., holding an infant) were evaluated more positively than were advertisements portraying agentic women (e.g., businesswomen).

Another tactic used in advertising is humor, which can increase an individual’s liking of a product based on positive association. For instance, Strick, Van Baaren, Holland, and Van Knippen (2009) found that products linked to a humorous cartoon were evaluated more positively in a subsequent assessment. Similarly, Krishnan and Chakravarti (2003) found that humor, even when completely irrelevant to the claim made by the advertisement, increased later recognition of the target product. Humor may elicit the attention of the viewer and increase positive emotions, thereby allowing them to look upon the product or company that is being advertised more favorably (Chan, 2011; Eisend, 2011). As Chan (2011) stipulated, humorous advertisements may have added persuasive effects, increasing the likelihood that a person will purchase a product or solicit a company’s services. Indeed, the lighthearted aspect of humor may allow for successful advertisement through the elicitation of positive emotions and product recognition.

Derogatory Humor

Offensive humor may be used as an attention-getting strategy, but often at a great price. In addition to circumscribing social roles for women, sexist stereotypes may further hostility and negative attitudes toward women. Indeed, derogatory humor may provide a venue in which people can more safely express their prejudices (Ford et al., 2008). For instance, Greenwood and Isbell (2002) found that men who responded positively to misogynistic jokes were especially likely to be hostile toward women (Greenwood & Isbell, 2002). The propensity to find humor in sexist jokes may vary by sex; in their study, Diaconu-Muresan and Stewart (2010) found that women exhibited more negative responses to sexist jokes than did men. Furthermore, participants who endorsed a feminist identity were less likely to respond positively to sexist jokes than were those low in feminist identity. Such results indicated that the propensity to find humor in these types of jokes may be mediated by gender and feminist identity.
Sexist humor has the potential to reinforce negative perceptions of women as well as internalized sexism. For instance, Ford et al. (2008) found that individuals exposed to sexist humor were less likely to contribute financially to a woman’s organization when exposed to a sexist joke; this suggests that exposure to sexist humor may detract from feminist causes. In addition, sexism and sexist jokes have been linked to rape myth acceptance (e.g., the belief that women who experience sexual assault are deserving of their trauma) among those who hear them (Romero-Sánchez et al., 2009). As the aforementioned evidence has indicated, sexist humor may reinforce hostility toward women, subsequently undermining their status, safety, and well-being.

**Justification of Sexism**

Of key interest to the present study was an understanding of how people respond to sexist humor in advertisements. We were particularly interested in the types of justifications and responses that people draw on when they view a sexist advertisement. Several theories seem particularly useful for understanding how sexism can be justified, minimized, and supported in current day society.

**Ambivalent sexism toward women.** Ambivalent sexism, introduced by Glick and Fiske (2001), emanates from the fact that men and women are believed to be dependent on one another despite the fact that women are believed to be the lesser sex. This form of sexism has two components: hostile and benevolent sexism. Hostile sexism toward women follows a pattern of contempt, criticizing women who defy traditional gender roles. In contrast, benevolent sexism adopts a seemingly positive perception of women by praising their communal nature and purity. Nevertheless, benevolent sexism reinforces women’s inferior status by emphasizing their dependent nature. As Glick and Fiske (2001) noted, hostile and benevolent sexism are positively correlated, and both correlate with support for gender inequality. Cross-culturally, men tend to endorse hostile sexism to a greater extent than do women. However, women are more likely to endorse benevolent sexism than hostile sexism, and the sex differences for endorsement of benevolent sexism are much smaller.

**Ambivalent sexism toward men.** In a complementary theory to that of ambivalent sexism toward women, Glick and Fiske (1999) proposed that ambivalently sexist attitudes toward men also serve to uphold traditional gender role stereotypes. Hostility toward men reflects women’s and some men’s anger that men hold greater power and dominate women. However, this hostility does not question the status difference, but rather asserts that men possess inherent traits such as aggression that result in their greater power and status. On the other hand, benevolent sexism toward men upholds expectations that women should take care of men because men are incapable of caring for themselves. Both forms of sexism reinforce gender inequality, providing justification for such inequality based on what are believed to be inherent differences between men and women (Glick & Fiske, 1999).

**Social Dominance Orientation (SDO).** Similarly, SDO, a theory popularized by Pratto, Sidanius, Stallworth, and Bertram (1994), justifies societal inequality by stating that such inequality is a natural and desirable aspect of social order and is impossible to avoid. SDO is associated with attitudes of racism and nationalism. In addition, those higher in SDO are less likely to endorse programs that advocate for the rights of women or LGBT-identified individuals. As Pratto et al. (1994) demonstrated, men tend to score higher on measures of SDO than women, an imbalance that may reflect the reluctance among high-status group members (e.g., men, the wealthy) to relinquish the benefits afforded to their positions of power.

**Cavalier humor beliefs.** Some may claim that offensive humor is only intended to evoke benign laughter. That is, such jokes are not believed to reflect nor promote prejudice. Such claims, labeled by Hodson, Rush, and MacInnis (2010) as cavalier humor beliefs, deny the potential for racist or sexist humor to perpetuate social inequality. However, cavalier humor beliefs are positively correlated with SDO as well as with racism (Hodson et al., 2010). These findings seem to indicate that a joke is not “just a joke,” and may allow joke tellers to disseminate racist and sexist ideas without any negative ramifications.

The aforementioned theories attempt to characterize relationships between social groups and how individuals attempt to create and reinforce a system of inequality. Based on this literature, it is likely that elements of each theory can be found in justifications for sexist humor and provide insight into how gender-based inequality is reinforced.

**Current Study**

Previous studies have indicated that advertisements often rely on stereotypes and humor to sell a
product (Strick et al., 2009). In addition, research has indicated that sexist humor is pervasive and prominent among individuals who hold implicitly prejudiced attitudes (Greenwood & Isbell, 2007). However, there has been a lack of research pertaining to justifications of sexist humor in the context of advertising and product promotion. The present study aimed to determine such factors by evaluating responses to a controversial label located within a pair of trousers distributed in the United Kingdom. This pants label included the washing instructions, “Give it to your woman . . . it’s her job,” reflecting the common gender role stereotype that women should be responsible for household chores. Of particular interest was how people responded to the stereotype embedded within these instructions. To examine reactions to the pants label, we conducted a content analysis of comments posted in response to news stories about the label in order to understand reactions to and justifications for a sexist marketing strategy.

Numerous major media sources in the United States including ABC®, CNN®, MSNBC®, MSN®, and AOL News® published a photo of the washing instructions on their respective websites, resulting in thousands of responses from readers. From each source, reader comments were examined and coded for their content.

Three primary questions guided the research process: foremost, to what extent do commentators endorse or respond positively to the sexist label, and to what extent do they dismiss or respond negatively to the label? Furthermore, what is the most common reaction to or justification for the label? In adopting the mirror perspective of gender role stereotypes in advertising, the present study aimed to understand the predominant attitudes regarding women, especially with respect to their domesticity, by evaluating responses to and justifications for the joke used in this pants label (Eisend, 2010).

Methods

Procedure

The comments selected for content analysis came from five mainstream U.S. news sources including ABC, CNN, MSNBC, AOL, and MSN. These news sources were selected because they covered the story of interest, have wide popularity, and most have been rated as relatively politically neutral (Groseclose & Milyo, 2005)1. AOL, ABC, and MSNBC each had relatively small numbers of respondents to the story on the article feedback site (AOL n = 19; ABC n = 36; MSNBC n = 31). MSN (n = 2,310) and CNN (n = 1,469) each had large numbers of responders. Thus, 135 MSN comments and 184 CNN comments were randomly selected from larger comment pools. Only the first published response was used for each user, and direct responses to other users were eliminated in order to maintain independence of observations. This resulted in the 406 comments used in the content analysis.

Content analysis coding. Each comment was analyzed and assigned to one or more of the following categories best representing its content: hostile sexism toward women, benevolent sexism toward women, hostile sexism toward men, benevolent sexism toward men, cavalier humor beliefs, expression of perceived humor, agreement with the statement’s observation of gender roles, SDO, denial of sexism altogether, and justification not otherwise specified (see Table 1 for summary of variables). Several comments were lengthy in response and had multiple components. These responses were deemed to fall into multiple categories, and were thus characterized by multiple codes. In the final dataset, 23.9% (n = 97) of the comments contained information that was either irrelevant to the label or too ambiguous to interpret and were thus labeled as uninterpretable.

In coding for ambivalent sexism toward women, hostile sexism toward women was inferred through hostility toward women, especially through claims that women were attempting to manipulate men. Furthermore, comments were labeled as indicative of hostile sexism when antagonistic remarks appeared to be intended to force women to conform to traditional gender roles. Sample comments included “Wow, this woman really needs some Prozac®” and “There appears to be allot [sic] of victimized drama queens posting.” On the other hand, comments were classified as benevolent sexism if they idealized women, upheld claims of inherent sex-based differences, or stressed interdependence between the sexes. Such responses resembled the following comment: “She washes my laundry all the time. She doesn’t want to mow the yard, wash the cars, take out the trash, clean the boat or the fish, etc...”

For ambivalent sexism toward men, comments that involved hostile sexism toward men reflected resentment of men’s superior social status as well

---

1 In the 10 years since Groseclose and Milyo (2005) was published, some of these news outlets may no longer be seen as politically neutral. Liberal-oriented readers may be more likely to perceive sexism in this particular comment, thereby skewing results. We thank a reviewer for noting this point.
It’s Just a Joke | Peters, Holmgreen, and Oswald

as references to negative male stereotypes (e.g., inability to take care of oneself or refusal to ask for directions). Sample comments included “If you can pull your pants up, then you should be able to wash them, pig,” and “All it says is that guys are clueless.” Benevolent sexism toward men was inferred by references to men’s dependent nature and women’s maternal duties to care for them (i.e., men’s incompetence in the domain of clothes-washing). Comments such as “I’ll do anything for my man” and “Sometimes us women are just naturally better at the laundry then men” were categorized in this way.

Comments that were characteristic of cavalier humor beliefs emphasized the positive aspects of joke-telling and minimized the potential social consequences of stereotyped humor. They tended to express sentiments such as “It’s just a joke” or “Lighten up.”

In contrast, comments coded as reflecting a perception of humor in the statement merely stated that “It’s funny,” or “LOL [common Internet shorthand for the phrase ‘laughing out loud’],” but did not explicitly tie the humor to gender-based identity or social disparity. Therefore, the perception of humor was distinguished from cavalier humor beliefs when the comment made no attempt to justify the perception of humor.

Comments that were coded as agreement provided a rationale for the accuracy of the pants label. Such responses were distinguished from other categories such as humor or SDO in that they basically reiterated the content of the label (i.e., “It is her job”) without further elaborating on their reasoning for endorsing the stereotype.

Comments were labeled as a denial of sexism if they attempted to justify the statement by denying the continued existence of sexist prejudice. Representative comments included “Women have it just as good as men these days” and “This isn’t 1950 anymore.” Such comments differed from cavalier humor beliefs in that they did not express a perception of humor but rather attempted to suppress the controversy surrounding the label by noting the significant social advances of women.

SDO comments included any sentiment that attempted to illustrate the inherent nature of sex differences. Such comments extended beyond mere agreement by providing explanations of gender roles emerging from biological differences or from long-standing, functional divisions of labor. Comments characteristic of SDO included “Am I the only one getting fed up with all the PC BS [politically correct bullsh*t] we’re subjected to when there are real issues to be dealt with” and “Men are men...and women are women. Let’s not try to be the same!”

Any comment that expressed resistance to the statement, for example, by stating that it was offensive or in poor taste, was classified as an objection to the label. For instance, one respondent stated, “Apparently it is still OK to make condescending [sic] jokes about women, and men’s ignorance and incompetence if that’s how you took it.”

Some responses did not fit into any of the categories listed above, and were thus labeled justification not otherwise specified. Finally, comments that were irrelevant or indiscernible were deemed uninterpretable.

To ensure coder reliability, a subsample of 40 items was coded by two of the authors with a percent agreement of 75% (Cohen’s kappa = .69). Because this level of agreement has been determined to be “substantial” in the literature, the remainder of the coding was completed by only one author (Landis & Koch, 1977).

**Results**

Overall, 7.6% (n = 31) of the comments expressed disagreement with or disapproval of the pants label. In contrast, 68.5% (n = 278) of the comments indicated some type of agreement with or justification for the label. A further 23.9% (n = 97) of the items were uninterpretable because of lack of relevance or because their content was too vague or unintelligible to be interpreted with confidence. A chi-square test of goodness-of-fit comparing these three code categories (i.e., agreement, disagreement, or uninterpretable) indicated that all codes were not equally responded to, $X^2(2, N = 406) = 241.70, p < .001$.

The most prevalent types of response are displayed in Table 2. The modal category of comment was cavalier humor beliefs, comprising 25.1% (n = 102) of the dataset. Hostile sexism toward women was the second most prevalent type of response, constituting 10.8% of comments (n = 44). Justification not otherwise specified characterized 10.1% of the comments (n = 41), and an equal number of individuals (10.8%) responded with mere humor (n = 31) or disagreement (n = 31). Hostile sexism toward men (n = 10) and denial of sexism categories (n = 10) each made up 2.5% of the dataset, respectively, and 2.0% of individuals responded with benevolent sexism toward men (n = 8). SDO characterized 1.7% of comments
Expression of offense

"Apparently, it is still OK
to make condescending [sic] jokes about women..."

Discussion

Results indicated that most respondents (68.5%) endorsed the sexist pants label in some way. In part, the prevalence of endorsement may be explained by cavalier humor beliefs. As the most common type of response, commentators tended to see the statement as “just a joke” and not a reflection of any greater injustice. There are two potential reasons why people might have reacted this way. First, the presence of humor may temporarily sanction offensive comments, allowing individuals to express latent values that might otherwise be rebuked. Indeed, Hodson et al. (2010) found a positive correlation between cavalier humor beliefs and racism, as well as a negative correlation between cavalier humor beliefs and social tolerance. Second, individuals who subscribe to cavalier humor beliefs may bear a certain degree of prejudice and treat offensive humor in a lighthearted manner in order to gain or maintain social acceptance.

On the other hand, the limited number of commentators (7.6%) who expressed objection to the label might have feared potential social ramifications of expressing displeasure with the joke. That is, readers who did find the label offensive might have feared ridicule or name-calling from the vast number of users who promoted the joke or alluded to sexist ideals. Consistent with the conclusions of Becker and Wright (2010), the presence of misogyny might have undermined the motivation for individuals to express distaste for the comment. In other words, the presence of hostile sexism might have silenced women who found offense, for fear that they might be lambasted by their peers or that they might even confirm the stereotype that women are “victimized drama queens.”

Although not nearly as common (7.6%), the humor category parallels that of cavalier humor in that it both endorses the joke and reinforces the stereotype by failing to challenge it. General humor was theoretically defined as a mere observation that the stereotype by failing to challenge it. General humor may also reflect the prejudicial values of respondents; as mentioned previously, Greenwood and Isbell (2002) found that sexist men were more likely to respond positively to derogatory humor. Given these findings, it is possible that a number of commentators embodied sexist values toward women prior to encountering this joke.

As expected, a number of comments (5.7%) expressed agreement, reflecting beliefs that doing laundry is an appropriate role for women. Laundry is a task that has historically been assigned to women, and several have indicated little change in this area over the past several decades (Lachance-Grzela & Bouchard, 2010; Ruppanner, 2010). Agreement with the statement reflects an endorsement of sexist cultural norms and fails to provide speculation as to why such norms exist, thereby potentially reinforcing the association of women with domesticity.

We anticipated that sexism directed toward both men and women would frequently be drawn upon to justify the sexist label. Interestingly, benevolent sexism toward women was not frequently expressed in the comments. One might have expected this to be a frequent category

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary of Content Analysis Codes</strong></td>
</tr>
<tr>
<td><strong>Code</strong></td>
</tr>
<tr>
<td>Hostile sexism toward women</td>
</tr>
<tr>
<td>Benevolent sexism toward women</td>
</tr>
<tr>
<td>Hostile sexism toward men</td>
</tr>
<tr>
<td>Cavalier humor beliefs</td>
</tr>
<tr>
<td>General humor</td>
</tr>
<tr>
<td>Agreement</td>
</tr>
<tr>
<td>Denial of sexism</td>
</tr>
<tr>
<td>Social dominance orientation</td>
</tr>
<tr>
<td>Objection</td>
</tr>
</tbody>
</table>


(n = 7), and only 0.5% of comments were characterized by benevolent sexism toward women (n = 2).
because “good traditional women” are revered for their excellent housekeeping skills. The lack of benevolently sexist comments may reflect reactions to the controversy inspired by this article; indeed, the story was brought to the attention of the media by a woman who was offended upon finding the label in her boyfriend’s pants. According to Glick and Fiske’s (2001) theory of ambivalent sexism, a woman’s rejection of the status quo might have been threatening to a sexist audience who perceived her outcry as an overreaction. Benevolent sexism is usually reserved for women who comply with traditional gender role norms, whereas hostile sexism is a more typical response to gender role defiance (Glick & Fiske, 2001).

A small percentage of individuals (2.5%) exhibited SDO. However, it is possible that such values were subsumed by the more relevant category of cavalier humor beliefs (given the fact that the topic of the story directly involved humor). According to Hodson and colleagues (2010), there is a correlation between cavalier humor beliefs and SDO. Therefore, it is possible that individuals who responded with cavalier humor beliefs were also prone to SDO, although they found expressions such as “It’s just a joke” to be more relevant than a rational and functional explanation of difference in the given situation.

Similarly, few responses (2.5%) were characteristic of the denial of sexism category. One of the functions of cavalier humor is to diminish the consequences of offensive joke telling. Therefore, those who responded with cavalier humor might have been implicitly denying sexism by claiming that the pants label was “just a joke” and not indicative of social inequality (Hodson et al., 2010).

Finally, benevolent (2.0%) and hostile sexism toward men (0.5%) were less frequent responses in our data. Both types of comments would necessarily have required respondents to make negative comments about men despite ambivalent sexism’s overriding function of upholding the imbalance of power between the sexes. Comments that referred to men’s incompetence in any domain might, then, seem unlikely in the context of this story. Alternatively, such comments might simply not have been made because they could have seemed irrelevant given the fact that women were the primary targets of this joke.

Limitations
The generalizability of these results might have been affected by the exclusion of various members of the population based on lack of Internet access. Those with access to the Internet may be more privileged, at least in terms of socioeconomic status, than those without access. Therefore, our results may be inflated in that they largely reflect the interests of high-status group members. Regardless, we determined that our dataset was highly representative of the population exposed to online media content and therefore generalizable to this particular subset of individuals.

The sources of data might have also affected the generalizability of these findings. We selected websites that had been previously identified as fairly neutral in terms of political affiliation. However, it is hard to determine the exact perspective of each news site, and the political affiliations of the readers would likely influence their reactions and justifications of the story. It would be interesting for future research to also integrate the political perspective of the website and commenter.

The present study might have benefited from the inclusion of demographic data such as sex of the commenters who participated. Our initial hope was to code comments for the sex of the writers. Unfortunately, such information was unavailable due to the low frequency of individuals who expressly stated their sex (n = 29). Future studies may explore the role that multiple intersecting facets of identity play in responses and reactions to sexist humor.

In addition, the fact that there were only two coders for the present study is a potential limitation. Because the interrater reliability on the initial

<table>
<thead>
<tr>
<th>Code</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavalier humor</td>
<td>109</td>
<td>25.1</td>
</tr>
<tr>
<td>Uninterpretable</td>
<td>97</td>
<td>23.9</td>
</tr>
<tr>
<td>Hostile sexism-men</td>
<td>44</td>
<td>10.8</td>
</tr>
<tr>
<td>Justification not otherwise specified</td>
<td>41</td>
<td>10.1</td>
</tr>
<tr>
<td>General humor</td>
<td>31</td>
<td>7.6</td>
</tr>
<tr>
<td>Objection</td>
<td>31</td>
<td>7.6</td>
</tr>
<tr>
<td>Agreement</td>
<td>23</td>
<td>5.7</td>
</tr>
<tr>
<td>Hostile sexism-women</td>
<td>10</td>
<td>2.5</td>
</tr>
<tr>
<td>Denial of sexism</td>
<td>10</td>
<td>2.5</td>
</tr>
<tr>
<td>Benevolent sexism-women</td>
<td>8</td>
<td>2.0</td>
</tr>
<tr>
<td>Social dominance</td>
<td>7</td>
<td>1.7</td>
</tr>
<tr>
<td>Benevolent sexism-men</td>
<td>2</td>
<td>0.5</td>
</tr>
</tbody>
</table>
subset of data was determined to be substantial, one coder proceeded with the coding of the rest of the comments. Given the established interrater reliability, we feel confident in the accuracy of the subsequent coding and the validity of the categories coded. Regardless, future researchers might consider using multiple independent coders for all comments to further ensure coding accuracy.

Despite the limitations of online-based content analysis, there is certainly value in such data. It is likely that the anonymity provided by such venues allows users to provide candid responses, and different results might be obtained in a laboratory due to social desirability demands. Thus, these types of comments provide a rich source of data that we encourage future researchers to consider using for qualitative research. Furthermore, qualitative coding programs such as Linguistic Inquiry and Word Count could help to further identify themes within the comments. Finally, future research could utilize experimental methods to determine exactly who accepts or rejects sexist humor and by exactly which mechanisms.

**Conclusion**

Many people claim that “a joke is just a joke,” yet evidence has suggested otherwise. As Hodson and colleagues (2010) indicated, individuals who view offensive humor with cavalier attitudes may, in fact, bear a certain degree of prejudice. The positive reaction elicited by this joke provided evidence of an unwillingness to reject the status quo. Furthermore, if Internet users respond publicly and positively to derogatory humor, this may increase the likelihood that companies will use similar marketing strategies in the future. Previous studies have provided evidence that exposure to derogatory humor may perpetuate negative attitudes toward historically oppressed groups (Ford et al., 2008; Greenwood & Isbell, 2002; Romero-Sánchez et al., 2009; Tiggemann & McGill, 2004). Exposure to prejudice such as sexism may be psychologically harmful. For example, Swim, Hyers, Cohen, and Ferguson (2001) found that women who experienced sexism in their everyday lives were more likely to report anger, anxiety, discomfort, depression, and low self-esteem. As this evidence indicated, a sexist joke may not be “just a joke.” Rather, it holds the potential to provoke attitudes of hostility and damage the psychological well-being of women. Determining the dynamics by which individuals justify and perpetuate sexist humor is an important first step in establishing efforts to counteract these processes. The present study contributed to an understanding of what types of comments may frequently emerge in the context of sexist humor, and future studies have the potential to further elucidate why these types of comments are used.

**References**


Author Note. Natalie Peters, Lucie Holmgreen, and Debra Oswald, Department of Psychology, Marquette University, WI. Lucie Holmgreen is now at Department of Behavioral Sciences, Rush University Medical Center. This research was supported by a Research Fellowship from the Women’s and Gender Studies Program at Marquette University. Correspondence concerning this article should be addressed to Natalie Peters, Milwaukee, WI. E-mail: nepeters@uwm.edu
Development of the capacity for self-control increases flexibility and adaptability, which may increase the ability to survive (Bushman, 2010). Parents often enroll children in martial arts programs to help them learn self-control and experience mastery. Conversely, theory and research on aggression (e.g., Bandura’s, 1971, Social Learning Theory) has suggested that teaching children to physically hurt others with their bodies (kicking, punching) may increase the likelihood that this behavior may be generalized to other situations, perhaps impulsively. If a child learns self-control strategies while learning martial arts, the child may be more able to manage aggressive motivation when provoked and may engage in less subsequent aggressive behavior. Whether a student of martial arts learns self-control regarding the use of physical force may depend on whether they are trained in traditional or modern martial arts.

**Traditional and Modern Martial Arts**

There are currently two main categories of martial arts: traditional and modern (Fuller, 1988).
Traditional martial arts instruction emphasizes psychological, spiritual, and nonaggressive aspects of the art, and modern martial arts tend to focus on competition and aggression (Fuller, 1988). Traditional martial arts training has been shown to reduce aggressive tendencies, and modern martial arts have been shown to increase the propensity to aggress in students who practice the training (Twemlow, Sacco, & Fanogy, 2008).

This pattern of results has been consistently found with aggressive youth. Trulson (1986) found decreases in self-reported aggression and anxiety, and increases in self-esteem in a group of aggressive male adolescents after a 6-month training program in traditional martial arts. In a similar study, Twemlow, Sacco, et al. (2008) found that these improvements remained at a 1-year follow-up. In contrast, a group that received modern martial arts training focused entirely on fighting skills actually increased in aggression compared to a control group that engaged in supervised physical activity that did not change significantly.

Different training schools may have different effects. Vertonghen and Theebom (2010) claimed that how the student is influenced by the martial art could be linked to a specific climate created by a traditional teaching of a martial art or a more modern competitive approach. These two types of approaches greatly influence how students respond to many other confrontations in their lives because martial arts teaches a person defense in case of physical assault, and also teaches coping mechanisms that can be utilized in situations that are not combative. Vertonghen and Theebom described the mastery climate and the performance climate. The mastery climate focuses on self-referenced improvement, and effort is rewarded; the performance climate encourages pupils to perform better than others. Therefore, it may be relevant to look at the motivational climate of different martial arts practices between different martial arts and also within the same martial art where different approaches are used (i.e., traditional versus modern).

Benefits of Traditional Martial Arts Training
Twemlow, Sacco, and colleagues (2008) posited several therapeutic effects of martial arts training in a traditional dojo, particularly for troubled youths. Training in traditional martial arts philosophy and codes of conduct may enhance character development and altruism. Traditional training can also foster positive attachments and develop skills fundamental to transform destructive aggressiveness into self-confidence. The practice of linking the mind with the body promotes self-awareness of physical movements and control of breathing and emotions such as fear and anger. This training may teach students how to stay calm when under great amounts of physical stress, improving the capacity to respond correctly. It is impossible to perfect a technique when a student is tense, out of control, angry, or inattentive because this leads to being off-balance and committing errors, which are often correctable in a balanced mindset. In a traditional dojo, students must accept that they need to show respect to the sensei in order to learn a desired skill. This change in thought is important because, instead of obeying the sensei merely because of their position of authority, the student develops a desire for self-discipline and then obeys in order to learn (Twemlow, Sacco, et al., 2008).

Twemlow, Biggs, et al. (2008) evaluated a traditional martial arts program implemented in three elementary schools named The Gentle Warrior Program, which was designed to reduce aggression in children. The sample consisted of 254 children in grades 3, 4, and 5, who participated in the Gentle Warrior Program as part of a larger school violence intervention. The results indicated that boys who participated in more martial arts training sessions reported a lower frequency of aggression and greater frequency of helpful “bystanding” (i.e., helpful behavior toward victims of bullying) over time, relative to boys with less frequent participation. The effect of participation on aggression was partially mediated by empathy, and empathy fully mediated the participation effect on helpful bystanding. No significant results were found for girls.

Twemlow, Biggs, et al. (2008) stated that the study provided preliminary support for the use of martial arts-based interventions to address bullying in schools for boys by teaching empathy, self-control, and peaceful strategies to resolve conflicts. The authors suggested that traditional martial arts-based training works because it allows students to role-play in a violent situation (getting grabbed or being punched) and decide to react nonviolently and not reciprocate the violence. The authors stated that specific skill training is more effective than didactic instruction. Practicing skills in a safe environment allows students to experience the kinds of emotions that they would face in a real-life confrontation and prepares them to react appropriately in that situation. Thus, mastery of...
a student’s body and development of associated self-control are critical aspects of traditional martial arts training.

**Hoi Jeon Moo Sool Internal Techniques**
The Korean martial art of Hoi Jeon Moo Sool is a traditional martial art that teaches students techniques of self-defense, hand-to-hand combat, kicking techniques, weapon (i.e., *external*) techniques, and breathing and meditation (i.e., *internal*) techniques (Myung, 2012). Internal techniques help the student focus, gain self-control, and decrease their heart rate through breathing techniques and predetermined movements. External techniques focus on conditioning the hands and feet to strike as well as conditioning the body to withstand blows. External techniques include strikes, throws, and self-defense moves such as escapes from grabs and defense against punches and kicks (H. Lozano, personal communication, August 16, 2012). Although internal and external techniques may blend with each other at times, it is appropriate to consider the techniques as two different categories.

To understand how internal and external training techniques can psychologically affect the individual who is practicing them, it is useful to consider them as environmental “inputs” that impact the individual. In the next section, we describe a current model of aggression that provides a framework about how training practices can influence the learner’s psychological experiences and behaviors.

**The General Aggression Model (GAM)**
The GAM (Anderson & Bushman, 2002) suggests that person and situation variables can increase hostile thoughts, feelings, and/or physiological arousal, which can then affect how a person perceives a situation and impulsive or thoughtful decisions about whether to aggress. If the person responds aggressively, this behavior may then act as a provocative stimulus toward another person, which can then affect another person’s behavioral reaction and recursively act as another situational input variable that again affects the original person (see Figure 1).

The present experiment tested the GAM. We hypothesized that participants who trained in external techniques would feel higher state hostility following training than those who trained in external and internal techniques. Further, participants training in only external techniques were anticipated to assign more punitive exercises (push-ups, leg-lifts, and sit-ups) to other students than those who trained in external and internal techniques. Using the GAM as a framework, a person variable in this experiment was how aggressive they tend to feel, think, and act across situations, which we will refer to as trait aggression. Two situation variables were included in the present experiment: internal techniques (e.g., breathing deeply) and external techniques (e.g., punching and kicking) used in traditional martial arts. Learning internal techniques was expected to decrease subsequent hostility and aggression more than those who did not practice internal techniques.

**Method**

**Participants**
Participants were students from two martial arts studios in San Antonio, TX. Fifty-two individuals participated. A 6-year-old participant was excluded from the analysis because she provided little data, and another participant was removed from the sample who did not complete many of the questionnaires and did not provide demographic information. Of the 50 resulting total participants, 35 were boys or men and 15 were girls or women. Participants’ ages ranged from 6 (one participant, to whom the questionnaires were read) to 58, with the mean age of 18.10 (SD = 12.63). Most participants identified their ethnicity as Hispanic (n = 22) and Mexican American (n = 16); six participants identified as European American, and six participants did not provide an ethnicity. The number of months that our participants studied martial arts ranged from 1 month to 168 months (M = 37.64, SD = 45.69).

**Procedure**
The Our Lady of the Lake University Institutional Review Board approved the experiment in May 2012. The first author received approval from the directors (senseis) of two dojos to conduct the experiment at their studios. Two weeks before the experiment, participant and guardian consent forms and a description of the general purpose and procedures were made available to the guardians of the dojo students. The first author invited students of each dojo to participate and explained the general purpose of the study. The students and parents were told that they could wait until the day of the experiment to sign the consent form, and we offered to answer any questions. The voluntary nature of the experiment was emphasized both
before and on the days that the experiment was conducted. All of the students (with their parents’ written consent) who attended lessons on the two days of the experiment, at the two dojos, elected to participate.

After all questions were answered and consent forms were collected, participants were asked to complete the Trait Aggression Questionnaire. Participants then received their regular external training, which included punching and kicking, and a series of exercises provided by their regular instructor for about 20 min. Through random assignment, half of the participants (the externally group; \( n = 25 \)) were moved to another room in the studio where they waited for 5 min for the experimenter. Participants in this condition then completed the State Hostility and Self-Discipline questionnaires and debriefing form. All questionnaires were presented orally to the 6-year-old participant.

The internal and external group (\( n = 25 \)) then received the internal training of deep, diaphragmatic breathing coupled with arm and leg movements. The regular instructor provided the internal training in one dojo, and the first author provided it in the other dojo. The content of the internal training provided was identical in both dojos. Participants were instructed to breathe in through their noses and allow their abdomens to extend while inhaling. They then exhaled through their mouths and contracted their stomach muscles while exhaling. After practicing this breathing, participants were shown how to move their bodies to assist the breathing techniques. Participants were led in inhaling quickly, then slowly, and breathing out forcefully (called a kiyup, which means yell in Korean). These internal techniques were explained and demonstrated two times. Then, the participants were led in performing the techniques for approximately 7 min. The internal techniques used in the present experiment were those taught to beginner students of Hoi Jeon Moo Sool. After completing the internal training for 7 min, the participants in this group completed the State Hostility and Self-Discipline questionnaires, and the debriefing form.

**Measures**

A demographic questionnaire was administered that included questions for ethnicity, age, and participant sex, as well as the number of months or years that the person had engaged in martial arts training.

**Trait aggression.** Trait aggression of the participants was measured with the 29-item Aggression Questionnaire (Buss & Perry, 1992). Participants rated the extent to which each statement was characteristic of them (1 = uncharacteristic to 7 = extremely characteristic). A sample item was “If somebody hits me, I hit back.” This measure contains four subscales: Physical Aggression, Verbal Aggression, Anger, and Hostility. Two items on the scale are phrased in a nonaggressive fashion and were reverse-scored, and mean variables were calculated for the whole scale, as well as for the four subscales. The number was found to be high in internal consistency in the present sample, Cronbach’s alpha = .92, as well as for each subscale (Physical Aggression (.76), Verbal Aggression (.78), Anger (.76), and Hostility (.86).

A median split variable was created for Trait Aggression in order to test for interaction effects with type of technique on the state hostility and aggression measures. All participants who reported a trait aggression score at or above the median of 2.95 on the Buss Perry Aggression Questionnaire were labeled high-trait aggression (\( n = 26 \)) and all who scored below were labeled low-trait aggression (\( n = 24 \)) for this variable.

**State hostility.** Following martial arts training (see Procedure section), current hostile affect was measured with the 35-item State Hostility Questionnaire (Anderson, Deuser, & DeNeve, 1995). Participants rated agreement with “I feel” statements such as “I feel outraged” using a Likert scale with the anchors of (1 = strongly disagree to 5 = strongly agree). Three participants voiced confusion regarding the item “I feel vexed.” Therefore,
that item was removed from the analyses. Eleven items were reverse worded, and thus were reverse scored. A mean state hostility variable was created. The Cronbach’s alpha for this measure was .91.

**Aggressive behavior.** We created a Self-Discipline Questionnaire that asked participants to imagine that they were the instructor of a martial arts class and that it was their job to help the student learn more self-discipline. They were then asked “how many push-ups would you have the student do?” Participants then circled a number between 1 and 50 push-ups, or wrote in the number if greater than 50. They then completed the same question for sit-ups and leg-lifts. This assignment of exercises (which can cause discomfort) was the operationalization of aggression. This is similar to measures of aggression that involved delivery of unpleasant stimuli in a laboratory context such as allocating the amount of hot sauce that a target is to ingest (White-Ajmani & Bursik, 2014).

**Design Overview**

The experiment utilized a 2 x 2 between-subjects design. The experimental independent variable was the type of technique taught with the levels of (a) external-only and (b) internal and external training. A correlational personality independent variable of trait aggression (with 4 subscale measures of Physical Aggression, Verbal Aggression, Anger, and Hostility) was also measured. The trait aggression measure was a continuous variable and was dichotomized into high and low categories for two-way Analyses of Variance (ANOVAs) to test for interaction effects with technique type. The dependent variables were (a) state hostility and (b) the number of push-ups, leg-lifts, and sit-ups allocated to another student in the class (as a measure of aggression).

**Results**

We conducted a frequency analysis to ensure that all data were entered correctly, and we looked for outliers using stem-and-leaf plots. No far outliers (Tukey, 1977) were found. Descriptive statistics for the measures were as follows: Trait Aggression ($M = 3.30; SD = 1.21$), State Hostility ($M = 2.01; SD = 0.64$), push-ups ($M = 35.82; SD = 18.94; range 10–100), sit-ups ($M = 41.26; SD = 21.01; range 20–100) and leg-lifts ($M = 33.38; SD = 16.65; range 10–99$).

To test how the continuous variables and subscales of the trait aggression measure were related to each other, correlational analyses were first conducted. Then the effects of the type of martial arts technique and the Trait Aggression x Technique Type interaction were tested with two-way ANOVAs.

**Correlational Analyses**

Bivariate correlations were conducted with all continuous variables including age, number of months in martial arts, mean responses on the Buss Perry Aggression Questionnaire, and its four subscales, state hostility mean ratings, and three aggression dependent variables of number of push-ups, leg-lifts, and sit-ups allocated to another student. All of the trait aggression subscales were significantly positively correlated with each other, all $p < .001$. Trait aggression was significantly positively correlated with state hostility, $r(48) = .65$, $p < .001$ (see Table 1). As predicted, trait aggression, state hostility, and all trait aggression subscales were each significantly positively correlated with at least two of the aggression measures of number of push-ups, sit-ups, and leg-lifts allocated to a peer student (see Table 2). Thus, participants with more aggressive personalities and those who felt higher state hostility inflicted more discomfort through exercise in a hypothetical situation than did those with less discomfort.

![Table 1](image)

| Correlations Between Trait Aggression, State Hostility, Months in Martial Arts (MA), and Push-Ups, Sit-Ups, and Leg-Lifts Allocated |
|------------------|------------------|------------------|------------------|------------------|
|                  | State hostility  | MA months        | Push-ups         | Sit-ups          | Leg-lifts        |
| Trait aggression | .65***           | -.16             | .44*             | .49***           | .33*             |
| State hostility  | —                | -.25             | .32**            | .32**            | .16              |
| MA months        | —                | —                | .32**            | -.15             | -.44***          |
| Push-ups         | —                | —                | —                | .53***           | .74***           |
| Sit-ups          | —                | —                | —                | —                | .48***           |

Note. *p < .05, **p < .01, ***p < .001.

![Table 2](image)

| Correlations Between Trait Aggression Subscales, State Hostility, and Push-Ups, Sit-Ups, and Leg-Lifts Allocated |
|------------------|------------------|------------------|------------------|------------------|
| Trait aggression subscale | State hostility  | Push-ups         | Sit-ups          | Leg-lifts        |
| Physical aggression    | .64***           | .37*             | .31*             | .19              |
| Verbal aggression      | .49***           | .22              | .44**            | .29              |
| Anger                 | .59***           | .32**            | .51***           | .27              |
| Hostility             | .45***           | .39*             | .38              | .38              |

Note. *p < .05, **p < .01, ***p < .001.
aggressive personalities and who felt lower state hostility. All of the allocated exercise aggression measures were also significantly positively correlated with each other, all ps < .01, which suggested that they measured similar constructs.

Interestingly, the number of months that participants had taken martial arts classes was significantly negatively correlated with assignment of push-ups, r(43) = -.32, p = .034, and leg-lifts, r(43) = -.44, p = .003. Perhaps students who had been in martial arts classes for a longer period of time better understood the pain associated with being assigned many push-ups and leg-lifts. Age was significantly negatively correlated with state hostility, r(47) = -.28, p = .048. Younger participants felt more hostile after the training than did older participants. Age was marginally negatively correlated with trait aggression, r(47) = -.24, p = .091, and the number of push-ups allocated, r(47) = -.25, p = .089. Age was not related to any other variables, all ps > .170.

Tests of Hypotheses: ANOVAs

Four two-way ANOVAs were conducted to test the main effects and interactions of type of technique and the dichotomous trait aggression variable on the dependent variables of state hostility and allocation of push-ups, sit-ups, and leg-lifts. Consistent with the correlational findings, a significant main effect of trait aggression was found for all dependent variables except for number of leg-lifts allocated. High-trait aggression was associated with more aggression than was low-trait aggression, all ps < .038. No significant interactions were found between trait aggression and technique type. However, the means were in the hypothesized direction. Participants in the external technique condition who were high in trait aggression reported more state hostility, and allocated more exercises to a peer student than did those lower in trait aggression. As expected, participants who had more aggressive personalities and who felt lower state hostility. All of the allocated exercise aggression measures were also significantly positively correlated with each other, all ps < .01, which suggested that they measured similar constructs.

A marginally significant technique type main effect was also found for allocated sit-ups, F(1, 46) = 3.19, p = .081, η² = .065. Participants with external-only training allocated significantly more sit-ups (M = 46.20, SD = 23.46) than those with internal/external training (M = 36.32, SD = 17.33; see Figure 4). No main effect of condition on allocated leg-lifts was found, p = .296.

Participant Sex and Ethnicity

No significant sex effects were found on any of the independent or dependent variables, all ps > .388. Meaningful comparisons could not be made across ethnicity because most participants self-identified as Hispanic or Mexican American (76%), and six participants (12%) did not indicate their ethnicity.

Discussion

Application of the GAM

As expected, participants who had more aggressive personalities were higher in trait aggression, reported more state hostility, and allocated more exercises to a peer student than did those lower in trait aggression. Consistent with the person route of the GAM, this finding provided further evidence that trait aggression predicts state hostility. It also extended the literature in demonstrating that trait aggression predicts justified aggression as measured by punitive exercise allocated to another person in order to build the target’s self-discipline.

![Figure 2: Effect of Type of Technique on State Hostility](image-url)
As hypothesized, participants who received internal training reported feeling slightly less hostile and allocated slightly fewer push-ups and sit-ups to another student after receiving the internal training than did participants who only received external training. This implied that martial arts training that includes internal breathing techniques may ameliorate the external training effects on aggressive motivation and behavior. Internal training may reduce actual and perceived physiological arousal and negative affect. According to the GAM, reducing physiological arousal and state hostility may increase the likelihood that a person will interpret an ambiguous action of another person as nonthreatening as opposed to someone who has not practiced internal techniques. Learning internal techniques may enable a person to make more thoughtful decisions when provoked because they may respond in a more relaxed and controlled fashion rather than in a fight or flight response to a potential threat. Consistent with Twemlow, Biggs, et al. (2008), integrating internal martial arts training into school physical education programs may therefore reduce bullying and overall aggression in students.

Limitations and Suggestions for Future Research
An important limitation that could be rectified in future research pertained to a procedural difference in the administration of the conditions. Participants exposed to both the external and the internal training experienced two training types and had two more minutes after the training before completing the dependent variables than did those who experienced the external-only training. It is possible that the slight time difference prior to completion of the dependent variables from the external training could have confounded the effects of the manipulation. Future experimental research could include a factorial design that tests the effects of internal and external training experienced separately and sequentially, compared to neither training type, on aggressive motivation and behavior. It would be important to ensure that the time between the training and when the dependent variables are administered is consistent. Physiological arousal following physical exertion associated with each training type should be strongly considered related to the timing of the subsequent completion of aggression measures. Arousal, particularly when a person is not aware of the arousal, has been associated with increased aggression (Zillmann & Bryant, 1974).

Our measurements were not based on real-world behavior but rather on self-reported behavioral intention measures that assessed what the participant would do as the dojo instructor. Another limitation was the small sample size, which might have reduced the likelihood of detecting interactions due to low statistical power. Furthermore, most of our participants were Hispanic, which makes the results difficult to generalize to members of other ethnic groups. Also, the specific internal breathing techniques used in the experiment were developed in the Hoi Jeon Moo Sool tradition so using techniques from other martial arts approaches may yield different results. The measures used were normed for college-aged students, thus the 6-year-old participant had a difficult time understanding the vocabulary and needed explanations of some words. Finally, participants from one dojo were less familiar with leg-lifts than

![Figure 3](image1.png)

**FIGURE 3**
Effect of Type of Technique on Push-Ups Allocated

![Figure 4](image2.png)

**FIGURE 4**
Effect of Type of Technique on Sit-Ups Allocated
Martial Arts Training and Aggression | Hernandez and Anderson

those from the other dojo.

Future research could test responses in real-world martial arts situations such as by having students spar after the external and internal/external training to score the number of aggressive acts they perform on another person. Alternatively, it would be helpful to observe how people trained in internal techniques react to a realistic aggressive act or stressful situation such as a possible provocation while waiting in line in the heat. More long-term studies of self-control (or aggressive) behavior in aggressive situations following frequent internal versus external-only martial arts training would be interesting. Internal training may decrease trait aggression and other hostile cognitive propensities such as the hostile attribution bias over time.

In conclusion, parents who seek martial arts training in order to develop enhanced self-control in their children are encouraged to enroll them in traditional martial arts programs. Most of these programs emphasize personal mastery through coordinated breathing and movement as well as ethical decision making before inflicting harm on another person. In contrast, many modern martial arts programs foster aggressive striking competition in which the winner “takes the other person out.”

References


Bushman, B. J. (2010). Effect of self-control on antisocial and prosocial behavior. Grant application submitted to the National Science Foundation.


Author Note. James Hernandez and Kathryn B. Anderson, Department of Psychology, Our Lady of the Lake University.

The authors are grateful to Brad Bushman, Henry Lozano, and Greg Borrego for their assistance with this project. Finally, special thanks are extended to the McNair Scholars Program at Our Lady of the Lake University for its support of this project.

Correspondence concerning this article should be addressed to Kathryn B. Anderson, Department of Psychology, Our Lady of the Lake University, San Antonio, TX 78207. E-mail: kbanderson@ollusa.edu
Four items are required for all submissions:

1) **Cover Letter**
   Include primary author's education status, manuscript originality statement, IRB approval

2) **Sponsoring Statement**
   Undergraduate first authors only

3) **Cover Page**
   Author names, school affiliation, and any author note

4) **Masked Manuscript**
   MS Word with all personal information removed

Simply register an account, then click Submit Manuscript:

1) **Upload Files**
   Files can be removed, replaced, or reorganized

2) **Enter Manuscript Information**
   E.g., title, abstract, authors, keywords, etc.

3) **Review Manuscript Material**
   Summary of all information/ files submitted

4) **Submit Manuscript**
   Receive email confirmation

New software benefits:
- Allows users to track their manuscripts' progress
- Inserts multiple files including cover letters, manuscripts, and figures
- Permits users to prioritize files and coauthors
- Checks for mistakes in the submission process and points out any errors
- Streamlines the process for authors and reviewers

* Psi Chi member ID number required

**Online Journal Submission Process**

All Psi Chi undergraduates, graduates, and faculty* are invited to submit their research to the *Psi Chi Journal of Psychological Research* through the new web based manuscript submission, tracking, and peer review software solution. Better than email submissions used in the past, this software allows users to create personal accounts to make the submission process more efficient.

**Register an account:** http://pcj.msubmit.net/cgi-bin/main.plx

**Tutorial videos:** http://www.ejpress.com/demos